PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C07H 21/02, C12Q 1/68, C12P 21/00, A01N 43/04, G06F 17/00 — A1

(11) International Publication Number: **WO 00/50436**

(43) International Publication Date: 31 August 2000 (31.08.00)

(21) International Application Number: PCT/US00/04606

(22) International Filing Date: 23 February 2000 (23.02.00)

(30) Priority Data:
60/121,314 23 February 1999 (23.02.99) US

(71) Applicant (*for all designated States except US*): GENAISSANCE PHARMACEUTICALS, INC. [US/US]; Five Science Park, New Haven, CT 06511 (US).

(71)(72) Applicants and Inventors: NANDABALAN, Krishnan [IN/US]; 228 Village Pond Road, Guilford, CT 06437 (US). SCHULZ, Vincent, P. [US/US]; 21 Old Farms Road, Madison, CT 06443 (US). STEPHENS, J., Claiborne [US/US]; 46 Crabapple Lane, Guilford, CT 06437 (US). CHEW, Anne [US/US]; Apartment 2, 837 Orange Street, New Haven, CT 06511 (US).

(74) Agent: HENDERSON, Melodie, W.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RECEPTOR ISOGENES: POLYMORPHISMS IN THE TISSUE NECROSIS FACTOR RECEPTOR

(57) Abstract

Polynucleotides comprising one or more of twelve novel single nucleotide polymorphisms in the human tumor necrosis factor receptor 1 (TNFR1) gene are described. Compositions and methods for detecting one or more of these polymorphisms are also disclosed. In addition, various genotypes and haplotypes for the TNFR1 gene that exist in the population are described.

(02(a) by another
less year

teaches 2, 3, 8-11, 13, 16

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

RECEPTOR ISOGENES:
POLYMORPHISMS IN THE TISSUE NECROSIS FACTOR RECEPTOR

This application is based on U.S. Provisional Application Serial No. 60/121,314 filed February 23, 1999.

Field of the Invention

This invention relates to variation in genes that encode pharmaceutically important proteins. In particular, this invention provides genetic variants of the human tumor necrosis factor receptor 1 (TNFR1) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

Background of the Invention

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a drug that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by a compound's activity at non-intended targets.

What this approach fails to consider, however, is that natural variability exists in any and every population with respect to a particular protein. A target protein currently used to screen drugs typically is expressed by a gene cloned from an individual who was arbitrarily selected. However, the nucleotide sequence of a particular gene may vary tremendously among individuals. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a target protein may be manifested as significant variation in expression of or in the structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in treatment of individuals with drugs whose design is based upon a single representative example of the target. For example, it is well-established that some classes of drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. In addition, variable information on the biological function or effects of a particular protein may be due to different scientists unknowingly studying different isoforms of the gene encoding the protein. Thus, information on the type and frequency of genomic variation that exists for pharmaceutically important proteins would be useful.

One possible target for the treatment of tumors, apoptosis-related and immunological disorders is tumor necrosis factor receptor 1 (TNFR1). The ligand for TNFR1 is tumor necrosis factor (TNF) which is secreted by macrophages, monocytes, neutrophils, T-cells and NK cells when they are stimulated. Two types of TNF are known, TNF alpha and TNF beta. TNFR1, also referred to as TNF receptor superfamily

member 1A (TNF-RSF1A), is a 55kDa protein that binds to both TNF alpha and beta (Hohmann et al., *J. Biol. Chem.* **264**:14927-14934, 1989). TNF exerts a spectrum of biological effects by binding to the TNFR1 receptor. Due to its cytotoxic and cytostatic effects, TNF can destroy the blood vessels in malignant tumors and can serve as an anti-tumor agent (Bruce et al., *Nature Med.* **2**: 788-794, 1996). TNF also mediates part of cell mediated immunity and confers resistance to infection caused by the facultative bacteria Listeria monocytogenes (Rothe et al., *Nature 364*; 798-802, 1993).

The C- terminal region of TNFR1 contains a death domain that interacts with MAP kinase-Activating Death Domain (MADD), a protein that acts as a mediator of the down stream effects of TNF signaling. MADD activates the MAP kinases and induces the phosphorylation of cytosolic phospholipase A2 (Schievella et al., *J Biol. Chem.* **272**: 12069-75, 1997). Autosomal dominant periodic fever syndrome, also known as TNF Receptor-Associated Periodic Syndromes (TRAPS), is characterized by episodes of fever and severe localized inflammation.

The TNFR1 gene in humans is located on chromosome 12p13 and the corresponding murine homolog is located on chromosome 6. The coding region and the 3'UTR of the TNFR1 gene are distributed over 10 exons (Fuchs et al., *Genomics* **13**: 219-224, 1992). The unprocessed precursor receptor is a glycosylated protein of 455 amino acids that contains a 29 amino acid signal sequence, an extracellular domain of 171 amino acids and a cytoplasmic domain of 221 amino acids (Loetscher et al., *Cell* **61**: 351-359). Although the full genomic sequence has not been published, a reference sequence for the TNFR1 gene comprises the non-contiguous sequences in Figure 1 (GenBank Accession No. X69810, Version X69810.1, GI:288493; SEQ ID NO:1) which contains exon 1, Figure 2 (GenBank Accession No. M75865, Version M75865.1, GI:339747; SEQ ID NO:2) which contains exons 2-5 and Figure 3 (GenBank Accession No. M75866, Version M75866.1, GI:339748; SEQ ID NO:3), which contains exons 6-10. Reference sequences for the TNFR1 coding sequence and amino acid sequence (Accession No. AAA61201.1; GI:339750) are shown in Figures 4 (SEQ ID NO:4) and 5 (SEQ ID NO:5) respectively.

A large number of single nucleotide polymorphisms in the TNFR1 coding sequence result in missense protein mutations. Six such missense mutations of this receptor have been studied in detail, of which five disrupt the conserved extracellular disulfide bonds (McDermott et al., *Cell* 97: 133-144, 1999). These mutations in the processed TNFR1 protein are Cys30Arg, Cys33Tyr, Thr50Met, Cys52Phe, Cys88Tyr and Cys88Arg. Membrane TNFR1 is regulated in part by metalloprotease-mediated cleavage where shedding of receptors followed by their clearance from the membrane takes place. Leukocytes bearing the Cys52Phe mutation showed increased levels of membrane receptor and diminished cleavage following stimulation. As a result, the down regulation of the membrane TNFR1 is impaired and the amount of soluble receptors in the cell decreases. This condition is manifested as an autoinflammatory syndrome (McDermott et al., *Cell* 97: 133-144, 1999).

Because of the potential for polymorphisms in the TNFR1 gene to affect tumor growth, apotosis, bacterial infection and disorders of the immune system, it would be useful to determine whether additional polymorphisms exist in the TNFR1 gene, as well as how such polymorphisms are combined in

different copies of the gene. Such information would be useful for studying the biological function of TNFR1 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

Summary of the Invention

Accordingly, the inventors herein have discovered twelve novel polymorphic sites in the TNFR1 gene. These polymorphic sites (PS) correspond to the nucleotide positions 201 (PS1), 230 (PS2), 845 (PS3), 873 (PS4) of Fig 1; nucleotide positions 481 (PS5), 526 (PS6), 839 (PS7), 880 (PS8), 971 (PS9), 1135 (PS10) of Fig 2; and nucleotide positions 162 (PS11) and 701 (PS12) of Fig 3. The polymorphisms at these sites are guanine or thymine at PS1, guanine or adenine at PS2, adenine or guanine at PS3, guanine or adenine at PS4, cytosine or thymine at PS5, cytosine or thymine at PS6, guanine or adenine at PS7, thymine or cytosine at PS8, cytosine or thymine at PS9, cytosine or thymine at PS10, guanine or adenine at PS11, and thymine or cytosine at PS12. In addition, the inventors have determined the identity of the alternative nucleotides present at these sites, in a human reference population of 112 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. It is believed that TNFR1-encoding polynucleotides containing one or more of the novel polymorphic sites reported herein will be useful in studying the expression and biological function of TNFR1, as well as in developing drugs targeting this protein. In addition, information on the combinations of polymorphisms in the TNFR1 gene may have diagnostic and forensic applications.

Thus, in one embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the TNFR1 gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS1, adenine at PS2, guanine at PS3, adenine at PS4, thymine at PS5, thymine at PS6, adenine at PS7, cytosine at PS8, thymine at PS9, thymine at PS10, adenine at PS11 and cytosine at PS12. A particularly preferred polymorphic variant is a naturally-occurring isoform (also referred to herein as an "isogene") of the TNFR1 gene. A TNFR1 isogene of the invention comprises guanine or thymine at PS1, guanine or adenine at PS2, adenine or guanine at PS3, guanine or adenine at PS4, cytosine or thymine at PS5, cytosine or thymine at PS6, guanine or adenine at PS7, thymine or cytosine at PS8, cytosine or thymine at PS10, guanine or adenine at PS11 and thymine or cytosine at PS12. The invention also provides a collection of TNFR1 isogenes, referred to herein as a TNFR1 genome anthology.

A TNFR1 isogene may be defined by the combination and order of these polymorphisms in the isogene, which is referred to herein as a TNFR1 haplotype. Thus, the invention also provides data on the number of different TNFR1 haplotypes found in the above four population groups. This haplotype data is useful in methods for deriving a TNFR1 haplotype from an individual's genotype for the TNFR1 gene and for determining an association between a TNFR1 haplotype and a particular trait.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a TNFR1 cDNA of a fragment thereof. The reference sequence comprises SEQ ID NO:4 (Fig. 4) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of guanine at a position corresponding to nucleotide 36, thymine at a position corresponding to nucleotide 224, thymine at a position corresponding to nucleotide 269, adenine at a position corresponding to nucleotide 362 and cytosine at a position corresponding to nucleotide 403 (see Fig 4). Polynucleotides complementary to these TNFR1 genomic variants are also provided by the invention.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express TNFR1 for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the TNFR1 protein. The reference amino acid sequence comprises SEQ ID NO:5 (Fig. 5) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of leucine at a position corresponding to amino acid position 75 (Leu75), methionine at a position corresponding to amino acid position 90 (Met90), glutamine at a position corresponding to amino acid position 121 (Gln 121) and histidine at a position corresponding to amino acid position 135 (His135) (see Fig. 5 and Table 2). A polymorphic variant of TNFR1 is useful in studying the effect the variant amino acid on the biological activity of TNFR1 as well as studying the binding affinity of candidate drugs targeting TNFR1 for the treatment of tumors, bacterial infection and disorders of the immune system and apoptosis-related disorders.

The present invention also provides antibodies that recognize and bind to one of the novel polymorphic TNFR1 protein variants described herein. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

In other embodiments, the invention provides methods, compositions, and kits for haplotyping and/or genotyping the TNFR1 gene in an individual. The methods involve identifying the nucleotide or nucleotide pair present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 in one or both copies of the TNFR1 gene from the individual. The compositions contain oligonucleotide probes and primers designed to specifically hybridize to one or more target regions containing, or that are adjacent to, a polymorphic site.

In yet another embodiment, the invention provides a method for identifying an association between a genotype or haplotype and a trait. In preferred embodiments, the trait is susceptibility to a disease or response to a drug. Such methods have applicability in developing diagnostic tests and therapeutic treatments for tumors, apotosis-related disorders, bacterial infection and disorders of the immune system.

The present invention also provides transgenic animals comprising one of the TNFR1 genomic

polymorphic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the TNFR1 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against TNFR1 protein, and for testing the efficacy of therapeutic agents and compounds for tumors, apotosis-related disorders, bacterial infection and disorders of the immune system in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the TNFR1 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the TNFR1 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing TNFR1 haplotypes organized according to their evolutionary relationships.

Brief Description of the Drawings

Figure 1 illustrates a partial reference sequence for the TNFR1 gene (contiguous lines; SEQ ID NO:1), with the underlines indicating start and stop codons, shading indicating the reference coding sequence, and bold nucleotides indicating the polymorphic sites and polymorphisms identified by Applicants in a reference population.

Figure 2 illustrates a partial reference sequence for the TNFR1 gene (contiguous lines; SEQ ID NO:2), with the underlines indicating start and stop codons, shading indicating the reference coding sequence, and bold nucleotides indicating the polymorphic sites and polymorphisms identified by Applicants in a reference population.

Figure 3 illustrates a partial reference sequence for the TNFR1 gene (contiguous lines; SEQ ID NO:3), with the underlines indicating start and stop codons, shading indicating the reference coding sequence, and bold nucleotides indicating the polymorphic sites and polymorphisms identified by Applicants in a reference population.

Figure 4 illustrates a reference sequence for the TNFR1 coding sequence (contiguous lines; SEQ ID NO:4), with underlines indicating the start and stop codons, and bold nucleotides indicating the polymorphic sites and polymorphisms identified by Applicants in a reference population.

Figure 5 illustrates a reference sequence for the TNFR1 protein (contiguous lines; SEQ ID NO:5), with the bold amino acids indicating the amino acid variations caused by the polymorphisms of Fig. 4.

Description of the Preferred Embodiments

The present invention is based on the discovery of novel variants of the TNFR1 gene. As described in more detail below, the inventors herein discovered twelve novel polymorphic sites by characterizing the TNFR1 gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 150 human individuals. The human individuals

included a reference population of 112 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (42 individuals), African descent (26 individuals) Asian (27 individuals) Hispanic/Latino (17 individuals). To the extent possible, the members of this reference population were organized into population subgroups by the self-identified ethnogeographic origin of their four grandparents as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent	Total	26
	Sierra Leone	1
Asian	Total	27
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian	Total	42
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino	Total	17
	Caribbean	7
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), five two- or three-generation Caucasian families from the CEPH-Utah cohort and one two-generation African-American family.

Using the TNFR1 genotypes identified in the Index Repository and the methodology described in the Examples below, the inventors herein also determined the haplotypes found on each chromosome for most human members of this repository. The TNFR1 genotypes and haplotypes found in the repository include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for studying population diversity, anthropological lineage, the significance of diversity and lineage at the phenotypic level, paternity testing, forensic applications, and for identifying associations between the genetic variation and a trait such as level of drug response or susceptibility to disease.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all known polymorphic sites in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the

present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

The inventors herein have discovered twelve novel polymorphic sites in the TNFR1 gene, which are referred to as of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 to designate the order in which they are located in the gene (see Figures. 1, 2 and 3 and Table 3 below).

Thus, in one embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the TNFR1 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant TNFR1 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12. Similarly, the nucleotide sequence of a variant fragment of the TNFR1 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of these polymorphic sites. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence (or other reported TNFR1 sequences) or to portions of the reference sequence (or other reported TNFR1 sequences), except for genotyping oligonucleotides as described below. The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence with SEQ ID NO:1, SEQ ID NO:2 and/or and SEQ ID NO:3. The polymorphism is selected from the group consisting of thymine at PS1, adenine at PS2, guanine at PS3, adenine at PS4, thymine at PS5, thymine at PS6, adenine at PS7, cytosine at PS8, thymine at PS9, thymine at PS10, adenine at PS11 and cytosine at PS12. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the TNFR1 gene which is defined by any one of haplotypes 1-17 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the TNFR1 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

TNFR1 isogenes may be isolated using any method that allows separation of the two "copies" of the TNFR1 gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and copending U.S. application Serial No. 08/987,966. Another method, which is described in copending U.S. Application Serial No. 08/987,966, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., Proc. Natl. Acad. Sci. 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 17 Nucleic Acids. Res. 8392, 1989; Ruaño et al., 19 Nucleic Acids Res. 6877-6882, 1991; Michalatos-Beloin et al., 24 Nucleic Acids Res. 4841-4843, 1996).

The invention also provides TNFR1 genome anthologies, which are collections of TNFR1 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A TNFR1 genome anthology may comprise individual TNFR1 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the TNFR1 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred TNFR1 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded TNFR1 protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant TNFR1 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 Science 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the

TNFR1 gene will produce TNFR1 mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of an TNFR1 cDNA comprising a nucleotide sequence which is a polymorphic variant of the TNFR1 reference coding sequence shown in Fig. 4. Thus, the invention also provides TNFR1 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:4, or its corresponding RNA sequence, except for having at least one variant nucleotide selected from the group consisting of guanine at a position corresponding to nucleotide 36, thymine at a position corresponding to nucleotide 224, thymine at a position corresponding to nucleotide 269, adenine at a position corresponding to nucleotide 362 and cytosine at a position corresponding to nucleotide 403. (see Fig 4). Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain at least one of these novel polymorphisms described herein. The location of a polymorphism in a variant mRNA, cDNA or fragment is identified by aligning its sequence with SEQ ID NO:4. The invention specifically excludes polynucleotides identical to previously identified and characterized TNFR1 cDNAs and fragments thereof except for genotyping oligonucleotides as described below. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

In describing the polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the TNFR1 gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the TNFR1 genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular TNFR1 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the TNFR1 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular TNFR1 isogene. Expression of a TNFR1 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions –10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and

Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of TNFR1 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of TNFR1 mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference TNFR1 amino acid sequence shown in Fig. 5 (SEQ ID NO:5). The location of a variant amino acid in a TNFR1 polypeptide or fragment of the invention is identified by aligning its sequence with Fig. 5. A TNFR1 protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:4 except for having one or more of leucine at a position corresponding to amino acid position 75 (Leu75), methionine at a position corresponding to amino acid position 90 (Met 90), glutamine at a position corresponding to amino acid position 121 (Gln 121) and histidine at a position corresponding to amino acid position 135 (His 135)(see Fig 5). The invention specifically excludes polypeptides consisting of amino acids identical to those previously identified for TNFR1, including SEQ ID NO:5, and previously described fragments thereof. TNFR1 protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:5 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a TNFR1 protein variant of the invention is encoded by an isogene defined by one of the haplotypes shown in Table 5. Such variants correspond to isoforms 1-4 of Table 2.

Table 2. Novel TNFR1 Protein Variants

ISOFORM NO:	Variant Amino Acids			
	75	90	121	135
1	L	T	R	Y
2	P	M	R	Y
3	P	T	Q	Y
4	P	T	R	H
5	P	T	Q	H
6	L	M	R	Y
7	L	T	Q	Y
8	P	M	R	H
9	L	T	R	H
10	P	M	Q	Y
11	L	M	Q	Y
12	P	M	Q	H
13	L	M	R	H
14	L	T	Q	H
15	L	M	Q	H

The invention also includes TNFR1 peptide variants, which are any fragments of a TNFR1 protein variant that contains one or more of the variant amino acid positions shown in Table 2. A TNFR1 peptide variant is at least 6 amino acids in length and is preferably any number between 6 an 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such TNFR1 peptide variants may be useful as antigens to generate antibodies specific for one of the above TNFR1 isoforms. In addition, the TNFR1 peptide variants may be useful in drug screening assays.

A TNFR1 variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant TNFR1 genomic and cDNA sequences as described above. Alternatively, the TNFR1 protein variant may be isolated from a biological sample of an individual having a TNFR1 isogene which encodes the variant protein. Where the sample contains two different TNFR1 isoforms (i.e., the individual has different TNFR1 isogenes), a particular TNFR1 isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular TNFR1 isoform but does not bind to the other TNFR1 isoform.

The expressed or isolated TNFR1 protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the TNFR1 protein as discussed further below. TNFR1 variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In "Current Protocols in Molecular Biology" John Wiley and Sons, New York, New York). In the case of immunoaffinity

chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant TNFR1 gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric TNFR1 protein. The non-TNFR1 portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the TNFR1 and non-TNFR1 portions so that the TNFR1 protein may be cleaved and purified away from the non-TNFR1 portion.

An additional embodiment of the invention relates to using a novel TNFR1 protein isoform or novel TNFR1 peptide variant in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known TNFR1 protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The TNFR1 protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to a TNFR1 variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the TNFR1 protein(s) of interest and then washed. Bound TNFR1 protein(s) are then detected using methods well-known in the art.

In another embodiment, the invention provides antibodies specific for and immunoreactive with the novel TNFR1 variant protein described herein. The antibodies may be either monoclonal or polyclonal in origin. The TNFR1 protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the CMA1 protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin ("Basic and Clinical Immunology", 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with a novel TNFR1 protein isoform is administered to an individual to neutralize activity of the TNFR1 isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with the novel TNFR1 protein isoform described herein may be used to immunoprecipitate the TNFR1 protein variant from solution as well as react with TNFR1 protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect TNFR1 protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with the novel TNFR1 protein variant described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the TNFR1 protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., *Principles and Practice of Immunoassay*, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; *Current Protocols in Molecular Biology*, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in *Methods in Immunodiagnosis*, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, *Methods in Immunology*, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., *Principles and Practice of Immunoassay*, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, *J. Clin. Chem. Clin. Biochem.*, 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in *Current Protocols in Molecular Biology, supra*.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, *Nature*, 256:495-497; Campbell "Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas", 1985, In: "Laboratory Techniques in Biochemistry and Molecular Biology," Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in *E. coli* is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, *Science*, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 *Proc. Natl. Acad. Sci.* 86;10029).

Effect(s) of the polymorphisms identified herein on expression of TNFR1 may be investigated by preparing recombinant cells and/or organisms, preferably recombinant animals, containing a polymorphic variant of the TNFR1 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into TNFR1 protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired TNFR1 isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene

will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the TNFR1 isogene is introduced into a cell in such a way that it recombines with the endogenous TNFR1 gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired TNFR1 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the TNFR1 isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the TNFR1 isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant organisms, i.e., transgenic animals, expressing a variant TNFR1 gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the TNFR1 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human TNFR1 isogene and producing human TNFR1 protein can be used as biological models for studying diseases related to abnormal TNFR1 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel TNFR1 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel TNFR1 isogenes; an antisense oligonucleotide directed against one of the novel TNFR1 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel TNFR1 isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders related to the expression

or function of a novel TNFR1 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Information on the identity of genotypes and haplotypes for the TNFR1 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of basic research and clinical applications. Thus, the invention also provides compositions and methods for detecting the novel TNFR1 polymorphisms identified herein.

The compositions comprise at least one TNFR1 genotyping oligonucleotide. In one embodiment, a TNFR1 genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical

synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a TNFR1 polynucleotide, i.e., a TNFR1 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-TNFR1 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the TNFR1 gene using the polymorphism information provided herein in conjunction with the known sequence information for the TNFR1 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruano et al., 87 Proc. Natl. Acad. Sci.

USA 6296-6300, 1990. Typically, an allele-specific oligonucleotide will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotide probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15 mer, the 8th or 9th position in a 16mer, the 10th or 11th position in a 20 mer). A preferred ASO probe for detecting TNFR1 gene polymorphisms comprises a nucleotide sequence selected from the group consisting of:

CAGACAG**G**TTCAGTT$^{3'}$ (SEQ ID NO:6) and its complement,

CAGACAG**T**TTCAGTT$^{3'}$ (SEQ ID NO:7) and its complement,

TTCATTT**G**TGTGTCC$^{3'}$ (SEQ ID NO:8) and its complement,

TTCATTT**A**TGTGTCC$^{3'}$ (SEQ ID NO:9) and its complement,

$^{5'}$TGCTGCC**A**CTGGTGA$^{3'}$ (SEQ ID NO:10) and its complement,

$^{5'}$TGCTGCC**G**CTGGTGA$^{3'}$ (SEQ ID NO:11) and its complement,

$^{5'}$GGGAAGA**G**TGGGCTG$^{3'}$ (SEQ ID NO:12) and its complement,

$^{5'}$GGGAAGAA**T**GGGCTG$^{3'}$ (SEQ ID NO:13) and its complement,

$^{5'}$C**C**AGGCCCGGGGCAG$^{3'}$ (SEQ ID NO:14) and its complement,

$^{5'}$C**C**AGGCC**T**GGGGCAG$^{3'}$ (SEQ ID NO:15) and its complement,

$^{5'}$TCCTTCA**C**CGCTTCA$^{3'}$ (SEQ ID NO:16) and its complement,

$^{5'}$TCCTTCA**T**CGCTTCA$^{3'}$ (SEQ ID NO:17) and its complement,

$^{5'}$G**T**GGACC**G**GGACACC$^{3'}$ (SEQ ID NO:18) and its complement,

$^{5'}$GTGGACC**A**GGACACC$^{3'}$ (SEQ ID NO:19) and its complement,

$^{5'}$CCGGC**A**TTATTGGAG$^{3'}$ (SEQ ID NO:20) and its complement,

$^{5'}$CCGGCATCATTGGAG$^{3'}$ (SEQ ID NO:21) and its complement,

$^{5'}$CTGAGGCCAAGCCCT$^{3'}$ (SEQ ID NO:22) and its complement,

$^{5'}$CTGAGGC**T**AAGCCCT$^{3'}$ (SEQ ID NO:23) and its complement,

$^{5'}$AGGCGCT**C**CTCCTTT$^{3'}$ (SEQ ID NO:24) and its complement,

$^{5'}$AGGCGCT**T**CTCCTTT$^{3'}$ (SEQ ID NO:25) and its complement,

$^{5'}$GAGGAGA**G**GTGACCT$^{3'}$ (SEQ ID NO:26) and its complement,

$^{5'}$GAGGAGAAGTGACCT$^{3'}$ (SEQ ID NO:27) and its complement,

$^{5'}$CTTTCTT**T**TTCCTCA$^{3'}$ (SEQ ID NO:28) and its complement and

$^{5'}$CTTTCTT**C**TTCCTCA$^{3'}$ (SEQ ID NO:29) and its complement.

An allele-specific oligonucleotide primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. Allele-specific oligonucleotide primers hybridizing to either the coding or noncoding strand are contemplated by the invention. A preferred ASO forward primer for detecting TNFR1 gene polymorphisms comprises a nucleotide sequence selected from the group consisting of:

5'CAGATCCAGACAGGT3' (SEQ ID NO:30),
5'CAGATCCAGACAGTT3' (SEQ ID NO:31),
5'GAGAAGTTCATTTGT3' (SEQ ID NO:32),
5'GAGAAGTTCATTTAT3' (SEQ ID NO:33),
5'ACCTGCTGCTGCCAC3' (SEQ ID NO:34),
5'ACCTGCTGCTGCCGC3' (SEQ ID NO:35),
5'GACAAAGGGAAGAGT3' (SEQ ID NO:36),
5'GACAAAGGGAAGAAT3' (SEQ ID NO:37),
5'GACTGTCCAGGCCCG3' (SEQ ID NO:38),
5'GACTGTCCAGGCCTG3' (SEQ ID NO:39),
5'AGCGGCTCCTTCACC3' (SEQ ID NO:40),
5'AGCGGCTCCTTCATC3' (SEQ ID NO:41),
5'TGCACAGTGGACCGG3' (SEQ ID NO:42),
5'TGCACAGTGGACCAG3' (SEQ ID NO:43),
5'CCAGTACCGGCATTA3' (SEQ ID NO:44),
5'CCAGTACCGGCATCA3' (SEQ ID NO:45),
5' AGTCTCCTGAGGCCA 3' (SEQ ID NO:46),
5'AGTCTCCTGAGGCTA3' (SEQ ID NO:47),
5'GGGTGCAGG CGCTCC3' (SEQ ID NO:48),
5'GGGTGCAGG CGCTTC3' (SEQ ID NO:49),
5'TCAGGTGAGGAGAGG3' (SEQ ID NO:50),
5'TCAGGTGAGGAGAAG3' (SEQ ID NO:51),
5'TGCTTTCTTTCTTTT3' (SEQ ID NO:52) and
5'TGCTTTCTTTCTTCT3' (SEQ ID NO:53).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein. In a preferred embodiment, the 3'-terminus of the genotyping oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site. A particularly preferred oligonucleotide primer for detecting TNFR1 gene polymorphisms by primer extension terminates in a nucleotide sequence selected from the group consisting of:

5' ATCCAGACAG 3' (SEQ ID NO:54); 5'CATAACTGAA3'(SEQ ID NO:55);
5'AAGTTCATTT3' (SEQ ID NO:56); 5'CTTGGACACA3'(SEQ ID NO:57);
5'TGCTGCTGCC3' (SEQ ID NO:58); 5'GTCTCACCAG3'(SEQ ID NO:59);
5'AAAGGGAAGA3' (SEQ ID NO:60); 5'CACCAGCCCA3' (SEQ ID NO:61);
5'TGTCCAGGCC 3' (SEQ ID NO:62); 5'ATCCTGCCCC3' (SEQ ID NO:63);

5'GGCTCCTTCA3' (SEQ ID NO:64); 5'TTCTGAAGCG3' (SEQ ID NO:65); 5'ACAGTGGACC3' (SEQ ID NO:66); 5'CACGGTGTCC3' (SEQ ID NO:67); 5'GTACCGGCAT3' (SEQ ID NO:68); 5'TCACTCCAAT3' (SEQ ID NO:69); 5'CTCCTGAGGC3' (SEQ ID NO:70); 5'GAGAGGGCTT3' (SEQ ID NO:71); 5'TGCAGGCGCT3' (SEQ ID NO:72); 5'GCTAAAGGAG3' (SEQ ID NO:73); 5'GGTGAGGAGA3' (SEQ ID NO:74); 5'ACCAGGTCAC3' (SEQ ID NO:75); 5'TTTCTTTCTT3' (SEQ ID NO:76); and 5'AACTGAGGAA3' (SEQ ID NO:77).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the TNFR1 gene in an individual. As used herein, the terms "TNFR1 genotype" and "TNFR1 haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the TNFR1 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid mixture comprising the two copies of the TNFR1 gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 in the two copies to assign a TNFR1 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12.

Typically, the nucleic acid mixture is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid mixture may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained

from an organ in which the TNFR1 gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ nontranscribed regions. If a TNFR1 gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid molecule containing only one of the two copies of the TNFR1 gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more of polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 in that copy to assign a TNFR1 haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the TNFR1 gene or fragment such as one of the methods described above for preparing TNFR1 isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two TNFR1 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional TNFR1 clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the TNFR1 gene in an individual. In a particularly preferred embodiment, the nucleotide at each of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 is identified.

In a preferred embodiment, a TNFR1 haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 in each copy of the TNFR1 that is gene present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 in each copy of the TNFR1 gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the TNFR1 gene, or fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous

at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., Proc. Natl. Acad. Sci. USA 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., Science 241:1077-1080, 1988). Oligonucleotides useful as primers or probes in such methods should specifically hybridize to a region of the nucleic acid that contains or is adjacent to the polymorphic site. Typically, the oligonucleotides are between 10 and 35 nucleotides in length and preferably, between 15 and 30 nucleotides in length. Most preferably, the oligonucleotides are 20 to 25 nucleotides long. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan.

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et L., Proc. Natl. Acad. Sci. USA 89:392-396, 1992.

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into

wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the TNFR1 gene of an individual may also be determined by hybridization of a nucleic sample containing one or both copies of the gene to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., Proc. Natl. Acad. Sci. USA 82:7575, 1985; Meyers et al., Science 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. Ann. Rev. Genet. 25:229-253 (1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., Genomics 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., Nucl. Acids Res. 18:2699-2706, 1990; Sheffield et al., Proc. Natl. Acad. Sci. USA 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. An other primer extension method is allele-specific PCR (Ruaño et al., Nucl. Acids Res. 17:8392, 1989; Ruaño et al., Nucl. Acids Res. 19, 6877-6882, 1991; WO 93/22456; Turki et al., J. Clin. Invest. 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In another aspect of the invention, an individual's TNFR1 haplotype pair is predicted from its TNFR1 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a TNFR1 genotype for the individual at two or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing TNFR1 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from

Caucasian, African American, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., Nucleic Acids Res. 24:4841-4843, 1996).

In one embodiment of this method for predicting a TNFR1 haplotype pair, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., Nucleic Acids Res. 24:4841-4843, 1996).

The invention also provides a method for determining the frequency of a TNFR1 genotype or

TNFR1 haplotype in a population. The method comprises determining the genotype or the haplotype pair for the TNFR1 gene that is present in each member of the population, wherein the genotype or haplotype comprises the nucleotide pair or nucleotide detected at one or more of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12 in the TNFR1 gene; and calculating the frequency any particular genotype or haplotype is found in the population. The population may be a reference population, a family population, a same sex population, a population group, a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for TNFR1 genotypes and/or haplotypes found in a reference population are used in a method for identifying an association between a trait and a TNFR1 genotype or a TNFR1 haplotype. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s) or haplotype(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s) or haplotype(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes and/or haplotypes observed in the populations are compared. If a particular genotype or haplotype for the TNFR1 gene is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that TNFR1 genotype or haplotype. Preferably, the TNFR1 genotype or haplotype being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 3 and 4, respectively, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting TNFR1 or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a TNFR1 genotype or haplotype, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be

obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the TNFR1 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and TNFR1 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their TNFR1 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the TNFR1 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention starts with a model of the form

$$r = r_0 + S \times d$$

where r is the response, r_0 is a constant called the "intercept", S is the slope and d is the dose. To determine the dose, the most-common and least common nucleotides at the polymorphic site are first defined. Then, for each individual in the trial population, one calculates a "dose" as the number of least-common nucleotides the individual has at the polymorphic site of interest. This value can be 0

(homozygous for the least-common nucleotide), 1 (heterozygous), or 2 (homozygous for the most common nucleotide). An individual's "response" is the value of the clinical measurement. Standard linear regression methods are then used to fit all the individuals' doses and responses to a single model (see e.g., L.D. Fisher and G. vanBelle, supra, Ch 9). The outputs of the regression calculation are the intercept r_0, the slope S, and the variance (which measures how well the data fits this simple linear model). The Students t-test value and the level of significance can then be calculated for each of the polymorphic sites.

A second method for finding correlations between TNFR1 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., supra Ch. 10), or other global or local optimization approaches (see discussion in Judson, supra) could also be used. As an example, a genetic algorithm approach is described herein. This method searches for optimal parameters or weights in linear or non-linear models connecting TNFR1 haplotype loci and clinical outcome. One model is of the form

$$C = C_0 + \sum_{\alpha}\left(\sum_i w_{i,\alpha} R_{i,\alpha} + \sum_i w'_{i,\alpha} L_{i,\alpha}\right) \qquad [1]$$

where C is the measured clinical outcome, i goes over all polymorphic sites, α over all candidate genes, C_0, $w_{i,\alpha}$ and $w'_{i,\alpha}$ are variable weight values, $R_{i,\alpha}$ is equal to 1 if site i in gene α in the first haplotype takes on the most common nucleotide and –1 if it takes on the less common nucleotide. $L_{i,\alpha}$ is the same as $R_{i,\alpha}$ except for the second haplotype. The constant term C_0 and the weights $w_{i,\alpha}$ and $w'_{i,\alpha}$ are varied by the genetic algorithm during a search process that minimizes the error between the measured value of C and the value calculated from Equation 1. Models other than the one given in Equation 1 can be readily incorporated by those skilled in the art for analyzing the clinical and polymorphism data. The genetic algorithm is especially suited for searching not only over the space of weights in a particular model but also over the space of possible models (Judson, supra).

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the TNFR1 gene. ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, supra, Ch. 10). These traits or variables are called the independent variables. To carry out ANOVA, the independent variable(s) are measured and individuals are placed into groups based on their values for these variables.

In this case, the independent variable(s) refers to the combination of polymorphisms present at a subset of the polymorphic sites, and thus, each group contains those individuals with a given genotype or haplotype pair. The variation in response within the groups and also the variation between groups is then measured. If the within-group response variation is large (people in a group have a wide range of responses) and the response variation between groups is small (the average responses for all groups are about the same) then it can be concluded that the independent variables used for the grouping are not causing or correlated with the response variable. For instance, if people are grouped by month of birth (which should have nothing to do with their response to a drug) the ANOVA calculation should show a low level of significance. However, if the response variation is larger between groups than within groups, the F-ratio (="between groups" divided by "within groups") is greater than one. Large values of the F-ratio indicate that the independent variable is causing or correlated with the response. The calculated F-ratio is preferably compared with the critical F-distribution value at whatever level of significance is of interest. If the F-ratio is greater than the Critical F-distribution value, then one may be confident that the individual's genotype or haplotype pair for this particular subset of polymorphic sites in the TNFR1 gene is at least partially responsible for, or is at least strongly correlated with the clinical response.

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of TNFR1 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the TNFR1 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the TNFR1 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying TNFR1 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the TNFR1 gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The TNFR1 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD ROM or on one or more other storage devices accessible by

the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

Example 1

This example illustrates examination of the TNFR1 gene for polymorphic sites from about 790 base pairs upstream of the ATG start site to about 860 base pairs upstream of the termination codon.

Amplification of Target Regions

Various regions of the TNFR1 gene were amplified using the following PCR primer pairs, with the indicated positions corresponding to GenBank Accession Nos. X69810, M75865 and M75866.

Accession No. X69180		
Fragment 1 (promoter)		
Forward primer		
14- 41	5'- CCAAGAAAGAGGGAGACTAGGAGGCTAG-3' (SEQ ID NO:78)	
Reverse primer		
complement of 328-301	5'- GATCACAGCAGTGAAAAGATCCAGAGTC-3' (SEQ ID NO:79)	
PCR product 315 nt.		
Fragment 2 (exon 1)		
Forward primer		
501-526	5'- GGCCTCCTCCTCCAGCTCTTCCTGTC-3' (SEQ ID NO:80)	
Reverse primer		
complement of 1139-1122	5'- ACATTCCCTTGGCCGCCC-3' (SEQ ID NO:81)	
PCR product	639 nt	
Accession No. M75865		
Fragment 3 (exon 2)		
Forward primer		
12-37	5'- CCCTAAGCTTCCCATCCCTCCCTCTC-3'(SEQ ID NO:82)	
Reverse primer		
complement of 280-254	5'-GCAGAGAAAGAAGCAGCACCCCAGACCTGAG-3'(SEQ IDNO:83)	
PCR product 269 nt		

Fragment 4 (exon 3)		
Forward primer		
367-393	5'- GAATCGGCCCTGGCTGTTGTCCCT AGC-3'(SEQ ID NO:84)	
Reverse primer		
complement of 702-675	5'- ACATCCATGCAGTGTCCCACCAAAAC-3'(SEQ ID NO:85)	
PCR product 336 nt		
Fragment 5 (exon 4)		
Forward primer		
724-750	5'- ACTTAGGGGTATGTCAGGAAGGGGATGCAGG-3' (SEQ ID NO:86)	
Reverse primer		
complement of 1046-1020	5'- GAATACAGGAGGGGGAAGGAAAGGAAG-3' (SEQ ID NO:87)	
PCR product 323 nt		
Fragment 6 (exon 5)		
Forward primer		
1081-1108	5'- CCCACCCTGTCCCTCGTCACTTCCTCTG-3'(SEQ ID NO:88)	
Reverse primer		
complement of 1351-1326	5'- ACTCTGGGGCATCCTGGCATCTGTTG-3' (SEQ ID NO:89)	
PCR product 244 nt		
Accession No. M75866		
Fragment 7 (exon 6)		
Forward primer		
9-35	5'- TGTGTGGTTGTTTTTCTGTGTTCCTC-3' (SEQ ID NO:90)	
Reverse primer		
complement of 241-217	5'- TGAGTCCTCAGTGCCCTTAACATTC-3' (SEQ ID NO:91)	
PCR product 143 nt		
Fragment 8 (exon 7)		
Forward primer		
219-245	5'- ATGTGTGGTTGT TTTTCTGTGT TCCTC-3' (SEQ ID NO:92)	
Reverse primer		
complement of 495-470	5'- TCACCTCCCTCCACACATGTCCATCG-3' (SEQ ID NO:93)	
PCR product 277 nt		
Fragment 9 (exon 8)		
Forward primer		
627-649	5'- AAGTCCCCACTGCCAGCTGAGTC-3'(SEQ ID NO:94)	
Reverse primer		
complement of 820-795	5'- ATGATTCCAGGGGATCTGAGCATTAG -3' (SEQ ID NO:95)	
PCR product 194 nt		

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 20 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 2 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl

Reverse Primer (10 μM) = 0.4 μl
Water =15.6μl

Amplification profile:
94°C - 2 min. 1 cycle

94°C - 30 sec.
70°C - 45 sec. } 10 cycles
72°C - 1 min.

94°C - 30 sec.
64°C - 45 sec. } 35 cycles
72°C - 1 min.

Sequencing of PCR Products

The PCR products were purified by Solid Phase Reversible Immobilization using the protocol developed by the Whitehead Genome Center. A detailed protocol can be found at http://www.genome.wi.mit.edu/sequencing/protocols/pure/SPRI_pcr.html.

Briefly, five μl of carboxyl coated magnetic beads (10 mg/ml) and 60 μl of HYB BUFFER (2.5M NaCl/20% PEG 8000) were added to each PCR reaction mixture (20 μl). The reaction mixture was mixed well and incubated at room temperature (RT) for 10 min. The microtitre plate was placed on a magnet for 2 min and the beads washed twice with 150 μl of 70% EtOH. The beads were air dried for 2 min and the DNA was eluted in 25 μl of distilled water and incubated at RT for 5 min. The beads were magnetically separated and the supernatant removed for testing and sequencing.

The purified PCR products were sequenced in both directions using the primer sets below:

Accession No. X69180
Fragment 1 (promoter)
Forward primer
39-61 5'- TAGTGAAGAACTCTGGAGTAAAG -3' (SEQ ID NO:96)
Reverse primer
complement of 310-287 5'- ATCCAGAGTCTGGGAGTCAAATTC-3'(SEQ ID NO:97)

Fragment 2 (exon 1)
Forward primer
531-554 5'- TGTTGCAACACT GCCTCACTCTTC -3'(SEQ ID NO:98)
Reverse primer
complement of 1022-997 5'-GGATGAATGGGGAACCCCACACTG-3'(SEQ ID NO:99)

Accession No. M75865
Fragment 3 (exon 2)
Forward primer
34-59 5'- TCTTGATGGTGTCTCCTCTATCTG-3' (SEQ ID NO:100)
Reverse primer
complement of 264-242 5'-CCAGACCTGAGGGCATTCACCGT-3' (SEQ ID NO:101)

Fragment 4 (exon 3)
Forward primer 3
90-413 5'- TAGCATGGGGCTCCTTCCTTGTGT-3 '(SEQ ID NO:102)
Reverse primer

complement of 680-656	5'- CAAAACACACACCTTCCTGCCCAC -3' (SEQ ID NO:103)
Fragment 5 (exon 4)	
Forward primer	
741-764	5'-GAAGGGGATGCAGGGACAGGAGGA-3' (SEQ ID NO:104)
Reverse primer	
complement of 1031-1007	5'- AAGGAAAGGAAGTGCCACCGCATG -3' (SEQ ID NO:105)
Fragment 6 (exon 5)	
Forward primer	
1091-1113	5'- CCCTCGTCACTTCCTCTGTCGTG -3' (SEQ ID NO:106)
Reverse primer	
complement of 1335-1311	5'- CATCTGTTGCCCAGCTAATGGTTC -3' (SEQ ID NO:107)
Accession No. M75866	
Fragment 7 (exon 6)	
Forward primer	
17-35	5'- TTGTTTTTCTGTGTTCCTCCAATG -3'(SEQ ID NO:108)
Reverse primer	
complement of 241-208	5'- ATGGGTGGGTAGGATGGACGGGTG -3' (SEQ ID NO:109)
Fragment 8 (exon 7)	
Forward primer	
226-249	5'- CACCCATCCATCTATCCCTGCGGC -3'(SEQ ID NO:110)
Reverse primer	
complement of 480-458	5'- CATGTCGATCGCACCCACCCATG -3'(SEQ ID NO:111)
Fragment 9 (exon 8)	
Forward primer	
638-660	5'- GCCAGCTGAGTCCAGGGTGCCAG-3' (SEQ ID NO:112)
Reverse primer	
complement of 806-784	5'- TCTGAGCATTAGGCAATTATAAG-3' (SEQ ID NO:113)

Analysis of Sequences for Polymorphic Sites

Sequences were analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., Nucleic Acids Res. 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the TNFR1 gene are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the TNFR1 Gene			
Polymorphic Site	**Nucleotide Position Number**	**Reference Allele**	**Variant Allele**
PS1	201 (Acc. No. X69810)	G	T
PS2	230 (Acc. No. X69810)	G	A
PS3	845 (Acc. No. X69810)	A	G
PS4	873 (Acc. No. X69810)	G	A
PS5	481 (Acc. No. M75865)	C	T
PS6	526 (Acc. No. M75865)	C	T
PS7	839 (Acc. No. M75865)	G	A
PS8	880 (Acc. No. M75865)	T	C
PS9	971 (Acc. No. M75865)	C	T
PS10	1135 (Acc.No. M75865)	C	T
PS11	162 (Acc. No. M75866)	G	A
PS12	701 (Acc. No. M75866)	T	C

Example 2

This example illustrates analysis of the TNFR1 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below.

Table 4. Observed Genotypes and Haplotype Pairs for TNFR1

Geno-type Number	Polymorphic Sites*												HAP Pair	
	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS 10	PS 11	PS 12		
1	G	A	A	G	C	C	G	T	C	C	A	C	1	1
2	G	A	A/G	G	C	C	G	T	C	C	G/A	T/C	1	2
3	G	G/A	A	G	C	C	G	T	C	C	A	C	1	3
4	G/T	A	A	G	C	C	G	T	C	C/T	A	T/C	1	5
5	G/T	A	A	G	C	C	G	T	C	C/T	A	C	1	6
6	G	A	A	G	C	C	G	T	C	C	A	T/C	1	9
7	G	G/A	A	G	C	C	G	T	C	C	A	T/C	1	11
8	G	G/A	A	G	C	C/T	G	T	C	C	A	T/C	1	17
9	G	A	G	G	C	C	G	T	C	C	G	T	2	2
10	G	G/A	A/G	G	C	C	G	T	C	C	G/A	T/C	2	3
11	G/T	A	A/G	G	C	C	G	T	C	C/T	A/G	T	2	5
12	G/T	A	A/G	G	C	C	G	T	C	C/T	G/A	T/C	2	6
13	G	A	A/G	G	C	C	G	T	C	C	G/A	T	2	9
14	G	G/A	A/G	G	C	C	G	T	C	C	A/G	T	2	11
15	G	A	G	G	C	C	G	T	C	C	G/A	T	2	12
16	G	G/A	A/G	G	C	C	G	T/C	C	C	G/A	T	2	14
17	G	G	A	G	C	C	G	T	C	C	G/A	C	3	4
18	G/T	G/A	A	G	C	C	G	T	C	C/T	A	T/C	3	5
19	G	G/A	A	G	C	C	G	T	C	C	A	T/C	3	9
20	G/T	G/A	A	G	C	C	G	T	C	C	A	C	3	13
21	G	G/A	A	G	C	C	G	T	C	C	G/A	T/C	4	9
22	G	G	A/G	G	C/T	C	G	T	C	C	G	T/C	4	10
23	T	A	A	G	C	C	G	T	C	T	A	T	5	5
24	T	A	A	G	C	C	G	T	C	T	A	T/C	5	6
25	T	A	A	G	C	C	G	T	C/T	T	A	T	5	7
26	T	A	A	G/A	C	C	G	T	C	T	A	T	5	8
27	G/T	A	A	G	C	C	G	T	C	C/T	A	T	5	9
28	T	A	A	G	C	C	G	T	C	C/T	A	T/C	5	13
29	T	A	A	G	C	C	G	T	C	C/T	G/A	T/C	5	15
30	G/T	A	A/G	G	C	C	G/A	T	C	C/T	G/A	T	5	16
31	G/T	A	A	G	C	C	G	T	C	C/T	A	T/C	6	9
32	G/T	A	A/G	G	C	C	G	T	C	C/T	A	T/C	6	12
33	G	G/A	A/G	G	C/T	C	G	T	C	C	G/A	T	9	10
34	G	G/A	A	G	C	C	G	T	C	C	A	T	9	11
35	G	G	A	G	C	C	G	T	C	C	A	T	11	11
36	G/T	G/A	A	G	C	C	G	T	C	C	A	T/C	11	13

*homozygous positions indicated by one nucleotide;
heterozygous positions indicated by two nucleotides.

TNFR1 haplotypes were derived or inferred from these genotypes using the following haplotype derivation protocol. First, the genotype data determined for the gene of interest are organized as a spreadsheet with one row per individual and one column per variable site. The pattern of segregation of each SNP in a genotype is followed in the family pedigrees in the Index Repository to authenticate it as a Mendelian variant. For instance, if both grandparents are homozygous G, a child or grand child should not be homozygous A.

From this list of validated genotypes, those that are homozygous at all sites or heterozygous at a single site were identified. The first such genotype is assigned as haplotype 1 (1 and 2 if a single site heterozygote), and all copies of this genotype are given the haplotype pair 1-1 (1-2 if a single site heterozygote). A list of known haplotypes for the population is created and updated as all homozygotes

and single site heterozygotes are assigned a haplotype pair. All such homozygous and single site heterozygous individuals are assigned a haplotype pair based on their haplotypes, adding new haplotypes to the list when they are first encountered.

Next the families are processed by invoking rules of Mendelian transmission of haplotypes. It is assumed that these are true families (e.g., no non-paternity) and that there was no mutation or recombination within the repository families. The authenticity of the CEPH family is well-known and has been established by the genotyping of thousands of genetic markers. Recombination is unlikely when the genomic region is less than a megabase, which corresponds to 1% recombination as a genomic average. With these assumptions, pedigrees are inspected and "worked" forward or backward. For instance, if a parent in a pedigree is a 1-1 homozygote, both grandparents and all children contain at least one copy of haplotype 1. Subtracting haplotype 1 from each of their genotypes reveals the other haplotype, which now can be used for further resolutions. Continuing this example, if the haplotype pair of one or more children is found to consist of haplotypes 1 and 5, the other parent, and at least one grandparent, must also have haplotype 5. Thus, haplotype 5 can be subtracted from their unphased genotypes to resolve their second haplotype, and subtract this new parental haplotype from the unresolved children and grandparent. In general, any homozygote in the pedigree is a reasonable place to start, but even "known" genotypes should be inspected for Mendelian transmission to detect potential errors. For instance, a 1-3 child is inconsistent with a 1-1 father and a 2-2 mother, even if all polymorphic sites were individually consistent with Mendelian transmission. Such cases can generally be resolved by identifying the source of error with respect to the rest of the pedigree -- grandparents and siblings in this case. Under the above assumptions, there can only be a maximum of four distinct genotypes among the children, representing the four combinations of parental haplotypes.

If no individuals in the pedigree are homozygous, a single-site heterozygote can usually be chosen to begin resolution of all haplotypes in the family. For instance, a 1-2 heterozygous child must have one parent with a 1 and the other with a 2. Although theoretically both parents could be consistent with either 1 or 2 (e.g., both parents are also 1-2's) this rarely happens in practice. As above, known haplotypes can be subtracted from the unphased genotypes to determine all haplotypes for the remainder of the family.

Although seldom observed this in practice, a family in which every individual is multiply heterozygous can be resolved by standard phase resolution in the transmission of haplotypes from parent to offspring, in which parental haplotypes are reconstructed from the pattern of occurrence of the individual variants among the offspring. Failing this, one can evaluate which of the known haplotypes potentially resolve the parents in the pedigree, and then work through each of these to see if Mendelian transmission is followed.

Because resolution of a multiply heterozygous genotype from an unrelated individual is completely unreliable without some idea of which haplotypes exist for the population, it is essential to directly determine at least some of the haplotypes from the population, e.g., by Mendelian analysis of genotypes in three-generation families and/or by a molecular haplotyping technique. Generally, only one

combination of known haplotypes is consistent with each individual's genotype, and that combination is inferred to be correct. Occasionally, only one known haplotype is consistent with one of these individuals, and so it is assumed present along with a new haplotype inferred from subtracting the known haplotype. Rarely, either no haplotypes or multiple pairs of haplotypes can be inferred, in which case the individual is not included in the haplotype data set.

Resolution proceeds among the unrelated individuals by using the growing list of known haplotypes to resolve phase as in the paragraph above. An additional rule is utilized in the case of polymorphisms that appear specific to a particular population group. If a new haplotype is required because of the occurrence of a previously unobserved variant, all individuals in that population group with the new variant are inspected to see if a single new haplotype will explain this variant. This is usually the case, and presumably represents an ethnogeographic-specific polymorphism.

Occasionally, a new SNP is observed in an individual that is multiply heterozygous. According to coalescence theory, which assumes there is no bias as to where a SNP can arise, this SNP should have arisen on the most frequent haplotype in the reference population because this should represent the oldest haplotype and thus the haplotype having the most time to evolve. However, since current haplotype frequencies may not actually correspond with haplotype age, a comparison is made between all the human haplotype(s) found in the repository and the chimpanzee haplotype(s)to determine which human haplotype(s) are consistent with the presumably more ancient haplotype(s) present in the chimpanzee. In addition, according to Hardy-Weinberg theory, a novel haplotype defined by a rare SNP should usually be found together with one of the more frequent haplotypes in a population group.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the seventeen human TNFR1 haplotypes shown in Table 5 below.

Table 5. Haplotypes Observed for TNFR1 Gene

Haplotype Number	Polymorphic Sites											
	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	PS11	PS12
1	G	A	A	G	C	C	G	T	C	C	A	C
2	G	A	G	G	C	C	G	T	C	C	G	T
3	G	G	A	G	C	C	G	T	C	C	A	C
4	G	G	A	G	C	C	G	T	C	C	G	C
5	T	A	A	G	C	C	G	T	C	T	A	T
6	T	A	A	G	C	C	G	T	C	T	A	C
7	T	A	A	G	C	C	G	T	T	T	A	T
8	T	A	A	A	C	C	G	T	C	T	A	T
9	G	A	A	G	C	C	G	T	C	C	A	T
10	G	G	G	G	T	C	G	T	C	C	G	T
11	G	G	A	G	C	C	G	T	C	C	A	T
12	G	A	G	G	C	C	G	T	C	C	A	T
13	T	A	A	G	C	C	G	T	C	C	A	C
14	G	G	A	G	C	C	G	C	C	C	A	T
15	T	A	A	G	C	C	G	T	C	C	G	C
16	G	A	G	G	C	C	A	T	C	C	G	T
17	G	G	A	G	C	T	G	T	C	C	A	T

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
   (a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for tumor necrosis factor receptor 1 (TNFR1) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS1, adenine at PS2, guanine at PS3, adenine at PS4, thymine at PS5, thymine at PS6, adenine at PS7, cytosine at PS8, thymine at PS9, thymine at PS10, adenine at PS11 and cytosine at PS12; and
   (b) a second nucleotide sequence which is complementary to the first nucleotide sequence.
2. The isolated polynucleotide of claim 1 which comprises a TNFR1 isogene.
3. The isolated polynucleotide of claim 1 which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.
4. A recombinant organism transformed or transfected with the isolated polynucleotide of claim 1, wherein the organism expresses an TNFR1 protein encoded by the first nucleotide sequence.
5. The recombinant organism of claim 4 which is a transgenic animal.
6. The isolated polynucleotide of claim 1, wherein the first nucleotide sequence is a polymorphic variant of a fragment of the TNFR1 gene, the fragment comprising one or more polymorphisms selected from the group consisting of thymine at PS1, adenine at PS2, guanine at PS3, adenine at PS4, thymine at PS5, thymine at PS6, adenine at PS7, cytosine at PS8, thymine at PS9, thymine at PS10, adenine at PS11 and cytosine at PS12.
7. An isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the TNFR1 cDNA or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:4 and the polymorphic variant comprises one or more polymorphisms selected from the group consisting of guanine at a position corresponding to nucleotide 36, thymine at a position corresponding to nucleotide 224, thymine at a position corresponding to nucleotide 269, adenine at a position corresponding to nucleotide 362 and cytosine at a position corresponding to nucleotide 403.
8. A recombinant organism transformed or transfected with the isolated polynucleotide of claim 7, wherein the organism expresses a tumor necrosis factor receptor 1 (TNFR1) protein encoded by the polymorphic variant sequence.
9. The recombinant organism of claim 8 which is a transgenic animal.
10. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the TNFR1 protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:5 and the polymorphic variant comprises one or more variant amino acids selected from the group consisting of leucine at a position corresponding to amino acid 75,

methionine at a position corresponding to amino acid 90, glutamine at a position corresponding to amino acid 121 and histidine at a position corresponding to amino acid 135.

11. An isolated antibody specific for and immunoreactive with the isolated polypeptide of claim 10.
12. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the tumor necrosis factor receptor 1 (TNFR1) gene at a polymorphic site selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12.
13. The composition of claim 12, wherein the genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele of the TNFR1 gene at a region containing the polymorphic site.
14. The composition of claim 13, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:6-29, the complements of SEQ ID NOS:6-29, and SEQ ID NOS:30-53.
15. The composition of claim 12, wherein the genotyping oligonucleotide terminates in a nucleotide sequence selected from the group consisting of SEQ ID NOS:54-77.
16. A method for genotyping the tumor necrosis factor receptor 1 (TNFR1) gene of an individual, comprising determining for the two copies of the TNFR1 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12.
17. The method of claim 16, wherein the determining step comprises:
    (a) isolating from the individual a nucleic acid mixture comprising both copies of the TNFR1 gene, or a fragment thereof, that are present in the individual;
    (b) amplifying from the nucleic acid mixture a target region containing at least one of the polymorphic sites;
    (c) hybridizing a genotyping oligonucleotide to one allele of the amplified target region, wherein the genotyping oligonucleotide terminates in a nucleotide sequence selected from the group consisting of SEQ ID NOS:54-77;
    (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the polymorphic site; and
    (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.
18. A method for haplotyping the tumor necrosis factor receptor 1 (TNFR1) gene of an individual which comprises determining, for one copy of the TNFR1 gene present in the individual, the identity of the nucleotide at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12.
19. The method of claim 18, wherein the determining step comprises

(a) isolating from the individual a nucleic acid molecule containing only one of the two copies of the TNFR1 gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid molecule a target region containing at least one of the polymorphic sites;

(c) hybridizing a genotyping oligonucleotide to one allele of the amplified target region, wherein the genotyping oligonucleotide terminates in a nucleotide sequence selected from the group consisting of SEQ ID NOS:54-77;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

20. A method for predicting a haplotype pair for the tumor necrosis factor receptor 1 (TNFR1) gene of an individual comprising:

(a) identifying an TNFR1genotype for the individual at two or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) accessing data containing the TNFR1haplotype pairs determined in a reference population; and

(d) assigning a haplotype pair to the individual that is consistent with the data.

21. A method for identifying an association between a trait and at least one genotype or haplotype of the tumor necrosis factor receptor 1 gene which comprises comparing the frequency of the genotype or haplotype in a population exhibiting the trait with the frequency of the genotype or haplotype in a reference population, wherein the genotype or haplotype comprises a nucleotide pair or nucleotide located at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11 and PS12, wherein a higher frequency of the genotype or haplotype in the trait population than in the reference population indicates the trait is associated with the genotype or haplotype.

22. The method of claim 21 wherein the haplotype is selected from haplotype numbers 1-17 shown in Table 5.

23. The method of claim 22, wherein the trait is a clinical response to a drug targeting TNFR1 .

24. A computer system for storing and analyzing polymorphism data for the tumor necrosis factor receptor 1 gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the genotypes and haplotype pairs shown in Table 4 and the haplotypes shown in Table 5.

25. A genome anthology for the tumor necrosis factor receptor 1 (TNFR1) gene which comprises TNFR1 isogenes defined by haplotypes 1-17 shown in Table 5.

26. A method for screening for drugs targeting a tumor necrosis factor receptor 1 (TNFR1) isoform which comprises contacting the TNFR1 isoform with a candidate agent and assaying for the binding activity, wherein the TNFR1 isoform comprises a polymorphic variant of a reference sequence for the TNFR1 protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:5 and the polymorphic variant comprises one or more variant amino acids selected from the group consisting of leucine at a position corresponding to amino acid 75, methionine at a position corresponding to amino acid 90, glutamine at a position corresponding to amino acid 121 and histidine at a position corresponding to amino acid 135.

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Nandabalan, Krishnan
Schulz, Vincent P
Stephens, J. Claiborne
Chew, Anne

<120> Drug Target Isogenes: Polymorphisms in the Tumor Necrosis Factor Receptor I Gene

<130> MWH0030PCT

<140> to be assigned
<141> 2000-02-23

<150> 60/121,314
<151> 1999-02-23

<160> 113

<170> PatentIn Ver. 2.1

<210> 1
<211> 1239
<212> DNA
<213> HOMO SAPIENS

<400> 1
gatcagtaaa ttcccaagaa agagggagac taggaggcta gtgaagaact ctggagtaaa 60
ggggaggatt actaagggac atggagtacc tatcatgtgt cggacgctta tctatatctc 120
tcccatctga acaaatcctt acaggaaccc caggagacag gttatctcca ctctgcaaat 180
tggaaaacag atccagacag gttcagttat gtgtctgaga agttcatttg tgtgtccaag 240
acacattctt agctaaaaag ctaagcattc tgaattggaa cccagagaat ttgactccca 300
gactctggat cttttcactg ctgtgatcca tctgggaaag gctagtgatg tgggcaaggg 360
cttattgccc cttggtgttt ggttgggagt ggtcggattg gtgggttggg ggcacaaggc 420
agccagatct gggactcctg tgcttgtgac tggactacaa agagttaaag aacgttgggc 480
ctcctcctcc cgcctcctgt ggcctcctcc tccagctctt cctgtcccgc tgttgcaaca 540
ctgcctcact cttcccctcc cactttctct cccctcctct ctgctttaat tttctcagaa 600
ttctctggac tgaggctcca gttctggcct ttgggttca agatcactgg gaccaggccg 660
tgatctctat gcccgagtct caaccctcaa ctgtcacccc aaggcacttg ggacgtcctg 720
gacagaccga gtcccgggaa gccccagcac tgccgctgcc acactgccct gagcccaaat 780
gggggagtga gaggccatag ctgtctggca tgggcctctc caccgtgcct gacctgctgc 840
tgccactggt gagaccaggg acaaagggaa gagtgggctg gtgggcgagg caccttccgg 900
ctggcgtggg ccctctccgg gaggggccg agcctctcct gcccgggcct ggtcctggcg 960
ccagtcaggc ctgcaggtcc taacctcagc cactgccagt gtgggggttcc ccattcatcc 1020
gcctttgga gtaggggctg cgctgaggca ggggaatggg agaagtttga aagggagaga 1080
gtaaaaggaa gccctggccc ctgacagcgg tggaagtttg tgggcggcca agggaatgtg 1140

ggcaggaggt aggcccaggg tggggcagat ttggcgggga aaagaaggga gtgggagtag 1200
gaagattagc gctcggggag tccagacggt tctgaattc 1239

<210> 2
<211> 1358
<212> DNA
<213> HOMO SAPIENS

<400> 2
agatctgcta cccctaagct tcccatccct ccctctcttg atggtgtctc ctctatctga 60
ttcttcccca ggtgctcctg gagctgttgg tgggaatata cccctcaggg gttattggac 120
tggtccctca cctaggggac agggagaaga gagatagtgt gtgtccccaa ggaaaatata 180
tccaccctca aaataattcg atttgctgta ccaagtgcca caaaggtagg ggcaagtgga 240
aacggtgaat gccctcaggt ctggggtgct gcttctttct ctgcttcttc cagttgttct 300
tccctaactt tgctgtctct cctgggctgg gattttctcc ctcctcctc tcctagagac 360
ttcagggaat cggccctggc tgttgtccct agcatggggc tccttccttg tgttctcacc 420
cgcagcctaa ctctgcggcc ccattcacag gaacctactt gtacaatgac tgtccaggcc 480
cggggcagga tacggactgc agggagtgtg agagcggctc cttcaccgct tcagaaaacc 540
acctcagaca ctgcctcagc tgctccaaat gccgaaaggg tgagtgtgca caggcaggag 600
agtcaggcgg gtcttgagtg gtgtgtgggt gcctgtctat gtgcaggctg gtgggtgtgg 660
gcaggaaggt gtgtgttttg gtgggacact gcatggatgt gagtgtgtat tacagagaca 720
cacacttagg ggtatgtcag gaaggggatg cagggacagg aggatgcagg actcataccc 780
catcttctcc cctcaccaga aatgggtcag gtggagatct cttcttgcac agtggaccgg 840
gacaccgtgt gtggctgcag gaagaaccag taccggcatt attggagtga aaacctttte 900
cagtgcttca attgcagcct ctgcctcaat gggaccgtgc acctctcctg tgagcgcagt 960
ctcctgaggc caagccctct ccccacccca ggggttggcc ccttccccat gcggtggcac 1020
ttcctttcct tccccctcct gtattctgtg ggtctgacaa ccaactcctc tctggccgcc 1080
cccaccctgt ccctcgtcac ttcctctgtc ctgtgggtcg ggggtgcagg cgctcctcct 1140
ttagctgtgc cgcacttctc cctacaggcc aggagaaaca gaacaccgtg tgcacctgcc 1200
atgcaggttt ctttctaaga gaaaacgagt gtgtctcctg tagtaagtga gtatctctga 1260
gagctgctgg gcactggatg gtggcatggg ttgggacggg tgactggtgg gaaccattag 1320
ctgggcaaca gatgccagga tgccccagag tgctcagg 1358

<210> 3
<211> 2254
<212> DNA
<213> HOMO SAPIENS

<400> 3
ccaacaatct gtgtggttgt ttttctgtgt tcctccaatg gtagggcctc tgttcaccag 60
tgccgtctct tcttttagct gtaagaaaag cctggagtgc acgaagttgt gcctacccca 120
gattgagaat gttaagggca ctgaggactc aggtgaggag aggtgacctg gtgcccatgc 180
tcacctgccc tctccctctt cttgccccca cccgtccatc catcccaccc atccatctat 240
ccctgcggcc cccctctgcc cgctcctctg accaacacct gctttgtctg caggcaccac 300
agtgctgttg cccctggtca ttttctttgg tctttgcctt ttatcctctc tcttcattgg 360
tttaatgtat cgctaccaac ggtggaagtc caagctctac tccattggtg agtgggggct 420

```
ttgggaggga gagggagctg gtgggggtga gggaggacat gggtgggtgc gatggacatg 480
tgtggaggga ggtgaggagt gtcccctcag ttcataccgc tggggactct gggcagaagg 540
tggccctgga tggctgggga gatgtcgagc tgcatcagta gctctctcgt ccctggggcc 600
acataggccc tgaggcatgt caccacaagt ccccactgcc agctgagtcc agggtgccag 660
ggctgagaga ggaagtgaaa tttatgatgc tttctttctt tttcctcagt ttgtgggaaa 720
tcgacacctg aaaaagaggt gagatgaaat gagagagtta ctcccaaatg tccctgacca 780
ttccttataa ttgcctaatg ctcagatccc ctggaatcat ccttcacttt ccgggggctc 840
gcctcattcc ctctaagtcc caacccccac gtagaataaa gagggccggg gctggttttc 900
gctgccgcac taatctgcgc caccttctct ctttcagggg gagcttgaag gaactactac 960
taagcccctg gccccaaacc caagcttcag tcccactcca ggcttcaccc ccaccctggg 1020
cttcagtccc gtgcccagtt ccaccttcac ctccagctcc acctataccc ccggtgactg 1080
tcccaacttt gcggctcccc gcagagaggt ggcaccaccc tatcaggggg ctgaccccat 1140
ccttgcgaca gccctcgcct ccgaccccat ccccaacccc cttcagaagt gggagggacag 1200
cgcccacaag ccacagagcc tagacagtgc gtttctcccg cggctggaga cgaggaggct 1260
gggggagggc cgggggagcg cgggaggcgt cagaggggac cacgagaggc ggagggcgcg 1320
ggatgcgggg cggggcctgg ggttgccgcc cgaggctcac cggcccgcgt ccccgcagct 1380
gatgaccccg cgacgctgta cgccgtggtg gagaacgtgc ccccgttgcg ctggaaggaa 1440
ttcgtgcggc gcctagggct gagcgaccac gagatcgatc ggctggagct gcagaacggg 1500
cgctgcctgc gcgaggcgca atacagcatg ctggcgacct ggaggcggcg cacgccgcgg 1560
cgcgaggcca cgctggagct gctgggacgc gtgctccgcg acatggacct gctgggctgc 1620
ctggaggaca tcgaggaggc gctttgcggc ccgccgccgc tcccgcccgc gcccagtctt 1680
ctcagatgag gctgcgcccc tgcggggcagc tctaaggacc gtcctgcgag atcgcctttcc 1740
aaccccactt tttctctgaa aggaggggtc ctgcaggggc aagcaggagc tagcagccgc 1800
ctacttggtg ctaacccctc gatgtacata gcttttctca gctgcctgcg cgccgccgac 1860
agtcagcgct gtgcgcgcgg agagaggtgc gccgtgggct caagagcctg agtgggtggt 1920
ttgcgaggat gagggacgct atgcctcatg cccgttttgg gtgtcctcac cagcaaggct 1980
gctcgggggc ccctggttcg tccctgagcc tttttcacag tgcataagca gttttttttg 2040
tttttgtttt gttttgtttt gtttttaaat caatcatgtt acactaatag aaacttggca 2100
ctcctgtgcc ctctgcctgg acaagcacat agcaagctga actgtcctaa ggcaggggcg 2160
agcacggaac aatggggcct tcagctggag ctgtggactt ttgtacatac actaaaattc 2220
tgaagttaaa gctctgctct tggagacagt ggct                          2254


<210> 4
<211> 1368
<212> DNA
<213> HOMO SAPIENS

<400> 4
atgggcctct ccaccgtgcc tgacctgctg ctgccgctgg tgctcctgga gctgttggtg 60
ggaatatacc cctcaggggt tattggactg gtccctcacc taggggacag ggagaagaga 120
gatagtgtgt gtccccaagg aaaatatatc caccctcaaa ataattcgat ttgctgtacc 180
aagtgccaca aaggaaccta cttgtacaat gactgtccag gcccggggca ggatacggac 240
tgcagggagt gtgagagcgg ctccttcacc gcttcagaaa accacctcag acactgcctc 300
agctgctcca aatgccgaaa ggaaatgggt caggtggaga tctcttcttg cacagtggac 360
cgggacaccg tgtgtggctg caggaagaac cagtaccggc attattggag tgaaaacctt 420
ttccagtgct tcaattgcag cctctgcctc aatgggaccg tgcacctctc ctgccaggag 480
aaacagaaca ccgtgtgcac ctgccatgca ggtttctttc taagagaaaa cgagtgtgtc 540
```

```
tcctgtagta actgtaagaa aagcctggag tgcacgaagt tgtgcctacc ccagattgag 600
aatgttaagg gcactgagga ctcaggcacc acagtgctgt tgcccctggt cattttcttt 660
ggtctttgcc ttttatccct cctcttcatt ggtttaatgt atcgctacca acggtggaag 720
tccaagctct actccattgt ttgtgggaaa tcgacacctg aaaaagaggg ggagcttgaa 780
ggaactacta ctaagcccct ggccccaaac ccaagcttca gtcccactcc aggcttcacc 840
cccaccctgg gcttcagtcc cgtgcccagt tccaccttca cctccagctc cacctatacc 900
cccggtgact gtcccaactt tgcggctccc cgcagagagg tggcaccacc ctatcagggg 960
gctgacccca tccttgcgac agcccctcgcc tccgacccca tccccaaccc ccttcagaag 1020
tgggaggaca gcgcccacaa gccacagagc ctagacactg atgaccccgc gtacgctgtac 1080
gccgtggtgg agaacgtgcc cccgttgcgc tggaaggaat tcgtgcggcg cctagggctg 1140
agcgaccacg agatcgatcg gctggagctg cagaacgggc gctgcctgcg cgaggcgcaa 1200
tacagcatgc tggcgacctg gaggcggcgc acgccgcggc gcgaggccac gctggagctg 1260
ctgggacgcg tgctccgcga catggacctg ctgggctgct tggaggacat cgaggaggcg 1320
ctttgcggcc ccgccgccct cccgcccgcg cccagtcttc tcagatga 1368
```

<210> 5
<211> 455
<212> PRT
<213> HOMO SAPIENS

<400> 5

```
Met Gly Leu Ser Thr Val Pro Asp Leu Leu Leu Pro Leu Val Leu Leu
  1               5                  10                  15

Glu Leu Leu Val Gly Ile Tyr Pro Ser Gly Val Ile Gly Leu Val Pro
            20                  25                  30

His Leu Gly Asp Arg Glu Lys Arg Asp Ser Val Cys Pro Gln Gly Lys
        35                  40                  45

Tyr Ile His Pro Gln Asn Asn Ser Ile Cys Cys Thr Lys Cys His Lys
    50                  55                  60

Gly Thr Tyr Leu Tyr Asn Asp Cys Pro Gly Pro Gly Gln Asp Thr Asp
65                  70                  75                  80

Cys Arg Glu Cys Glu Ser Gly Ser Phe Thr Ala Ser Glu Asn His Leu
                85                  90                  95

Arg His Cys Leu Ser Cys Ser Lys Cys Arg Lys Glu Met Gly Gln Val
            100                 105                 110

Glu Ile Ser Ser Cys Thr Val Asp Arg Asp Thr Val Cys Gly Cys Arg
        115                 120                 125

Lys Asn Gln Tyr Arg His Tyr Trp Ser Glu Asn Leu Phe Gln Cys Phe
    130                 135                 140
```

```
Asn Cys Ser Leu Cys Leu Asn Gly Thr Val His Leu Ser Cys Gln Glu
145                 150                 155                 160

Lys Gln Asn Thr Val Cys Thr Cys His Ala Gly Phe Phe Leu Arg Glu
                165                 170                 175

Asn Glu Cys Val Ser Cys Ser Asn Cys Lys Lys Ser Leu Glu Cys Thr
            180                 185                 190

Lys Leu Cys Leu Pro Gln Ile Glu Asn Val Lys Gly Thr Glu Asp Ser
        195                 200                 205

Gly Thr Thr Val Leu Leu Pro Leu Val Ile Phe Phe Gly Leu Cys Leu
    210                 215                 220

Leu Ser Leu Leu Phe Ile Gly Leu Met Tyr Arg Tyr Gln Arg Trp Lys
225                 230                 235                 240

Ser Lys Leu Tyr Ser Ile Val Cys Gly Lys Ser Thr Pro Glu Lys Glu
                245                 250                 255

Gly Glu Leu Glu Gly Thr Thr Thr Lys Pro Leu Ala Pro Asn Pro Ser
            260                 265                 270

Phe Ser Pro Thr Pro Gly Phe Thr Pro Thr Leu Gly Phe Ser Pro Val
        275                 280                 285

Pro Ser Ser Thr Phe Thr Ser Ser Ser Thr Tyr Thr Pro Gly Asp Cys
    290                 295                 300

Pro Asn Phe Ala Ala Pro Arg Arg Glu Val Ala Pro Pro Tyr Gln Gly
305                 310                 315                 320

Ala Asp Pro Ile Leu Ala Thr Ala Leu Ala Ser Asp Pro Ile Pro Asn
                325                 330                 335

Pro Leu Gln Lys Trp Glu Asp Ser Ala His Lys Pro Gln Ser Leu Asp
            340                 345                 350

Thr Asp Asp Pro Ala Thr Leu Tyr Ala Val Val Glu Asn Val Pro Pro
        355                 360                 365

Leu Arg Trp Lys Glu Phe Val Arg Arg Leu Gly Leu Ser Asp His Glu
    370                 375                 380

Ile Asp Arg Leu Glu Leu Gln Asn Gly Arg Cys Leu Arg Glu Ala Gln
385                 390                 395                 400
```

Tyr Ser Met Leu Ala Thr Trp Arg Arg Arg Thr Pro Arg Arg Glu Ala
405 410 415

Thr Leu Glu Leu Leu Gly Arg Val Leu Arg Asp Met Asp Leu Leu Gly
420 425 430

Cys Leu Glu Asp Ile Glu Glu Ala Leu Cys Gly Pro Ala Ala Leu Pro
435 440 445

Pro Ala Pro Ser Leu Leu Arg
450 455

<210> 6
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 6
cagacaggtt cagtt 15

<210> 7
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 7
cagacagttt cagtt 15

<210> 8
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 8
ttcatttgtg tgtcc 15

<210> 9
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 9
ttcatttatg tgtcc 15

<210> 10
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 10
tgctgccact ggtga 15

<210> 11
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 11
tgctgccgct ggtga 15

<210> 12
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 12
gggaagagtg ggctg 15

<210> 13
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 13
gggaagaatg ggctg 15

<210> 14
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 14
ccaggcccgg ggcag 15

<210> 15

<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 15
ccaggcctgg ggcag 15

<210> 16
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 16
tccttcaccg cttca 15

<210> 17
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 17
tccttcatcg cttca 15

<210> 18
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 18
gtggaccggg acacc 15

<210> 19
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 19
gtggaccagg acacc 15

<210> 20
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 20
ccggcattat tggag 15

<210> 21
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 21
ccggcatcat tggag 15

<210> 22
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 22
ctgaggccaa gccct 15

<210> 23
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 23
ctgaggctaa gccct 15

<210> 24
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 24
aggcgctcct cctttt 15

<210> 25
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 25
aggcgcttct cctttt 15

<210> 26
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 26
gaggagaggt gacct 15

<210> 27
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 27
gaggagaagt gacct 15

<210> 28
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 28
ctttcttttt cctca 15

<210> 29
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 29
ctttcttctt cctca 15

<210> 30
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 30
cagatccaga caggt 15

<210> 31

<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 31
cagatccaga cagtt 15

<210> 32
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 32
gagaagttca tttgt 15

<210> 33
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 33
gagaagttca tttat 15

<210> 34
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 34
acctgctgct gccac 15

<210> 35
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 35
acctgctgct gccgc 15

<210> 36
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 36
gacaaaggga agagt 15

<210> 37
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 37
gacaaaggga agaat 15

<210> 38
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 38
gactgtccag gcccg 15

<210> 39
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 39
gactgtccag gcctg 15

<210> 40
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 40
agcggctcct tcacc 15

<210> 41
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 41
agcggctcct tcatc 15

<210> 42
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 42
tgcacagtgg accgg 15

<210> 43
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 43
tgcacagtgg accag 15

<210> 44
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 44
ccagtaccgg catta 15

<210> 45
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 45
ccagtaccgg catca 15

<210> 46
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 46
agtctcctga ggcca 15

<210> 47

<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 47
agtctcctga ggcta 15

<210> 48
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 48
gggtgcaggc gctcc 15

<210> 49
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 49
gggtgcaggc gcttc 15

<210> 50
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 50
tcaggtgagg agagg 15

<210> 51
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 51
tcaggtgagg agaag 15

<210> 52
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 52
tgctttcttt ctttt 15

<210> 53
<211> 15
<212> DNA
<213> HOMO SAPIENS

<400> 53
tgctttcttt cttct 15

<210> 54
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 54
atccagacag 10

<210> 55
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 55
cataactgaa 10

<210> 56
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 56
aagttcattt 10

<210> 57
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 57
cttggacaca 10

<210> 58
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 58
tgctgctgcc 10

<210> 59
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 59
gtctcaccag 10

<210> 60
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 60
aaagggaaga 10

<210> 61
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 61
caccagccca 10

<210> 62
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 62
tgtccaggcc 10

<210> 63

<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 63
atcctgcccc 10

<210> 64
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 64
ggctccttca 10

<210> 65
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 65
ttctgaagcg 10

<210> 66
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 66
acagtggacc 10

<210> 67
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 67
cacggtgtcc 10

<210> 68
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 68
gtaccggcat 10

<210> 69
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 69
tcactccaat 10

<210> 70
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 70
ctcctgaggc 10

<210> 71
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 71
gagagggctt 10

<210> 72
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 72
tgcaggcgct 10

<210> 73
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 73
gctaaaggag 10

<210> 74
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 74
ggtgaggaga 10

<210> 75
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 75
accaggtcac 10

<210> 76
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 76
tttctttctt 10

<210> 77
<211> 10
<212> DNA
<213> HOMO SAPIENS

<400> 77
aactgaggaa 10

<210> 78
<211> 28
<212> DNA
<213> HOMO SAPIENS

<400> 78
ccaagaaaga gggagactag gaggctag 28

<210> 79

<211> 28
<212> DNA
<213> HOMO SAPIENS

<400> 79
gatcacagca gtgaaaagat ccagagtc 28

<210> 80
<211> 26
<212> DNA
<213> HOMO SAPIENS

<400> 80
ggcctcctcc tccagctctt cctgtc 26

<210> 81
<211> 18
<212> DNA
<213> HOMO SAPIENS

<400> 81
acattccctt ggccgccc 18

<210> 82
<211> 26
<212> DNA
<213> HOMO SAPIENS

<400> 82
ccctaagctt cccatcccta cctctc 26

<210> 83
<211> 31
<212> DNA
<213> HOMO SAPIENS

<400> 83
gcagagaaag aagcagcacc ccagacctga g 31

<210> 84
<211> 27
<212> DNA
<213> HOMO SAPIENS

<400> 84
gaatcggccc tggctgttgt ccctagc 27

<210> 85
<211> 26
<212> DNA
<213> HOMO SAPIENS

<400> 85
acatccatgc agtgtcccac caaaac 26

<210> 86
<211> 31
<212> DNA
<213> HOMO SAPIENS

<400> 86
acttaggggt atgtcaggaa ggggatgcag g 31

<210> 87
<211> 27
<212> DNA
<213> HOMO SAPIENS

<400> 87
gaatacagga gggggaagga aaggaag 27

<210> 88
<211> 28
<212> DNA
<213> HOMO SAPIENS

<400> 88
cccaccctgt ccctcgtcac ttcctctg 28

<210> 89
<211> 26
<212> DNA
<213> HOMO SAPIENS

<400> 89
actctggggc atcctggcat ctgttg 26

<210> 90
<211> 26
<212> DNA
<213> HOMO SAPIENS

<400> 90
tgtgtggttg tttttctgtg ttcctc 26

<210> 91
<211> 25
<212> DNA
<213> HOMO SAPIENS

<400> 91
tgagtcctca gtgcccttaa cattc 25

<210> 92
<211> 27
<212> DNA
<213> HOMO SAPIENS

<400> 92
atgtgtggtt gtttttctgt gttcctc 27

<210> 93
<211> 26
<212> DNA
<213> HOMO SAPIENS

<400> 93
tcacctccct ccacacatgt ccatcg 26

<210> 94
<211> 23
<212> DNA
<213> HOMO SAPIENS

<400> 94
aagtccccac tgccagctga gtc 23

<210> 95

<211> 26
<212> DNA
<213> HOMO SAPIENS

<400> 95
atgattccag gggatctgag cattag 26

<210> 96
<211> 23
<212> DNA
<213> HOMO SAPIENS

<400> 96
tagtgaagaa ctctggagta aag 23

<210> 97
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 97
atccagagtc tgggagtcaa attc 24

<210> 98
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 98
tgttgcaaca ctgcctcact cttc 24

<210> 99
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 99
ggatgaatgg ggaaccccac actg 24

<210> 100
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 100
tcttgatggt gtctcctcta tctg 24

<210> 101
<211> 23
<212> DNA
<213> HOMO SAPIENS

<400> 101
ccagacctga gggcattcac cgt 23

<210> 102
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 102
tagcatgggg ctccttcctt gtgt 24

<210> 103
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 103
caaaacacac accttcctgc ccac 24

<210> 104
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 104
gaaggggatg cagggacagg agga 24

<210> 105
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 105
aaggaaagga agtgccaccg catg 24

<210> 106
<211> 23
<212> DNA
<213> HOMO SAPIENS

<400> 106
ccctcgtcac ttcctctgtc ctg 23

<210> 107
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 107
catctgttgc ccagctaatg gttc 24

<210> 108
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 108
ttgtttttct gtgttcctcc aatg 24

<210> 109
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 109
atgggtgggt aggatggacg ggtg 24

<210> 110
<211> 24
<212> DNA
<213> HOMO SAPIENS

<400> 110
cacccatcca tctatccctg cggc 24

<210> 111

<211> 23
<212> DNA
<213> HOMO SAPIENS

<400> 111
catgtcgatc gcacccaccc atg 23

<210> 112
<211> 23
<212> DNA
<213> HOMO SAPIENS

<400> 112
gccagctgag tccagggtgc cag 23

<210> 113
<211> 23
<212> DNA
<213> HOMO SAPIENS

<400> 113
tctgagcatt aggcaattat aag 23

POLYMORPHISMS IN THE HOMO SAPIENS TUMOR NECROSIS FACTOR RECEPTOR 1(TNFR1) GENE

GenBank Accession No. X69810 GI:288493

```
GATCAGTAAA TTCCCAAGAA AGAGGGAGAC TAGGAGGCTA GTGAAGAACT
CTGGAGTAAA GGGGAGGATT ACTAAGGGAC ATGGAGTACC TATCATGTGT        100
CGGACGCTTA TCTATATCTC TCCCATCTGA ACAAATCCTT ACAGGAACCC
CAGGAGACAG GTTATCTCCA CTCTGCAAAT TGGAAAACAG ATCCAGACAG        200
GTTCAGTTAT GTGTCTGAGA AGTTCATTTG TGTGTCCAAG ACACATTCTT
T                              A                  *
AGCTAAAAAG CTAAGCATTC TGAATTGGAA CCCAGAGAAT TTGACTCCCA        300
GACTCTGGAT CTTTTCACTG CTGTGATCCA TCTGGGAAAG GCTAGTGATG
TGGGCAAGGG CTTATTGCCC CTTGGTGTTT GGTTGGGAGT GGTCGGATTG        400
GTGGGTTGGG GGCACAAGGC AGCCAGATCT GGGACTCCTG TGCTTGTGAC
TGGACTACAA AGAGTTAAAG AACGTTGGGC CTCCTCCTCC CGCCTCCTGT        500
GGCCTCCTCC TCCAGCTCTT CCTGTCCCGC TGTTGCAACA CTGCCTCACT
CTTCCCCTCC CACCTTCTCT CCCCTCCTCT CTGCTTTAAT TTTCTCAGAA        600
TTCTCTGGAC TGAGGCTCCA GTTCTGGCCT TTGGGGTTCA AGATCACTGG
GACCAGGCCG TGATCTCTAT GCCCGAGTCT CAACCCTCAA CTGTCACCCC        700
AAGGCACTTG GGACGTCCTG GACAGACCGA GTCCCGGGAA GCCCCAGCAC
TGCCGCTGCC ACACTGCCCT GAGCCCAAAT GGGGGAGTGA GAGGCCATAG        800
CTGTCTGGCA TGGGCCTCTC CACCGTGCCT GACCTGCTGC TGCCACTGGT
                                                   G
GAGACCAGGG ACAAAGGGAA GAGTGGGCTG GTGGGCGAGG CACCTTCCGG        900
                   A
CTGGCGTGGG CCCTCTCCGG GAGGGGGCCG AGCCTCTCCT GCCCGGGCCT
GGTCCTGGCG CCAGTCAGGC CTGCAGGTCC TAACCTCAGC CACTGCCAGT       1000
GTGGGGTTCC CCATTCATCC GCCTTTTGGA GTAGGGGCTG CGCTGAGGCA
GGGGAATGGG AGAAGTTTGA AAGGGAGAGA GTAAAAGGAA GCCCTGGCCC       1100
CTGACAGCGG TGGAAGTTTG TGGGCGGCCA AGGGAATGTG GGCAGGAGGT
AGGCCCAGGG TGGGGCAGAT TTGGCGGGGA AAAGAAGGGA GTGGGAGTAG       1200
GAAGATTAGC GCTCGGGGAG TCCAGACGGT TCTGAATTC
```

FIGURE 1

POLYMORPHISMS IN THE HOMO SAPIENS TUMOR NECROSIS FACTOR RECEPTOR 1(TNFR1) GENE

GenBank Accession No. M75865 GI:339747

```
AGATCTGCTA CCCCTAAGCT TCCCATCCCT CCCTCTCTTG ATGGTGTCTC
CTCTATCTGA TTCTTCCCCA GGTGCTCCTG GAGCTGTTGG TGGCAATATA      100
CCCCTCAGGG GTTATTGGAC TGGTCCCTCA CCTAGGGGAC AGGGAGAAGA
GAGATAGTGT GTGTCCCCAA GGATAAATATA TCCACCCTCA AAATAATTCG     200
ATTTGCTGTA CCAAGTGCCA CAAAGGTAGG GGCAAGTGGA AACGGTGAAT
GCCCTCAGGT CTGGGGTGCT GCTTCTTTCT CTGCTTCTTC CAGTTGTTCT     300
TCCCTAACTT TGCTGTCTCT CCTGGGCTGG GATTTTCTCC CTCCCTCCTC
TCCTAGAGAC TTCAGGGAAT CGGCCCTGGC TGTTGTCCCT AGCATGGGGC     400
TCCTTCCTTG TGTTCTCACC CGCAGCCTAA CTCTGCGGCC CCATTCACAG
GAACCTACTT GTACAATGAC TGTCCAGGCC CGGGGCAGGA TACGGACTGC     500
                          T
AGGGAGTGTG AGAGCGGCTC CTTCACCGCT TCAGAAAACC ACCTCAGACA
                      T
CTGCCTCAGC TGCTCCAAAT GCCGAAAGGG TGAGTGTGCA CAGGCAGGAG     600
AGTCAGGCGG GTCTTGAGTG GTGTGTGGGT GCCTGTCTAT GTGCAGGCTG
GTGGGTGTGG GCAGGAAGGT GTGTGTTTTG GTGGGACACT GCATGGATGT     700
GAGTGTGTAT TACAGAGACA CACACTTAGG GGTATGTCAG GAAGGGGATG
CAGGGACAGG AGGATGCAGG ACTCATACCC CATCTTCTCC CCTCACCAGA     800
AATGGGTCAG GTGGAGATCT CTTCTTGCAC AGTGGACCGG GACACCGTGT
                                  A
GTGGCTGCAG GAAGAACCAG TACCGGCATT ATTGGAGTGA AAACCTTTTC     900
                        C
CAGTGCTTCA ATTGCAGCCT CTGCCTCAAT GGGACCGTGC ACCTCTCCTG
TGAGCGCAGT CTCCTGAGGC CAAGCCCTCT CCCCACCCCA GGGGTTGGCC    1000
                  T
CCTTCCCCAT GCGGTGGCAC TTCCTTTCCT TCCCCCTCCT GTATTCTGTG
GGTCTGACAA CCAACTCCTC TCTGGCCGCC CCCACCCTGT CCCTCGTCAC    1100
TTCCTCTGTC CTGTGGGGTG GGGGTGCAGG CGCTCCTCCT TTAGCTGTGC
                              T
CGCACTTCTC CCTACAGGCC AGGAGAAACA GAACACCGTG TGCACCTGCC    1200
ATGCAGGTTT CTTTCTAAGA GAAAACGAGT GTGTCTCCTG TAGTAAGTGA
GTATCTCTGA GAGCTGCTGG GCACTGGATG GTGGCATGGG TTGGGACGGG    1300
TGACTGGTGG GAACCATTAG CTGGGCAACA GATGCCAGGA TGCCCCAGAG
TGCTCAGG
```

FIGURE 2

POLYMORPHISMS IN THE HOMO SAPIENS TUMOR NECROSIS FACTOR RECEPTOR (TNFR1) GENE

GenBank Accession No. M75866 GI:339748

```
CCAACAATCT GTGTGGTTGT TTTTCTGTGT TCCTCCAATG GTAGGGCCTC
TGTTCACCAG TGCCGTCTCT TCTTTTAGCT GTAAGAAAAG CCTGGAGTGC                100
ACGAAGTTGT GCCTACCCCA GATTGAGAAT GTTAAGGGCA CTGAGGACTC
AGGTGAGGAG AGGTGACCTG GTGCCCATGC TCACCTGCCC TCTCCCTCTT                200
      A
CTTGCCCCCA CCCGTCCAT GTGTGGTTGT TTTTCTGTGT TCCTCCGGCC
CCCCTCTGCC CGCTCCTCTG ACCAACACCT GCTTTGTCTG CAGGCACCAC                300
AGTGCTGTTG CCCCTGGTCA TTTTCTTTGC TCTTTGCCTT TTATCCCTCC
TCTTCATTGG TTTAATGTAT CGCTACCAAC GGTGGAAGTC CAAGCTCTAC                400
TCCATTGGTG AGTGGGGGCT TTGGGAGGGA GAGGGAGCTG GTGGGGGTGA
GGGAGGACAT GGGTGGGTGC GATGGACATG TGTGGAGGGA GGTGAGGAGT                500
GTCCCCTCAG TTCATACCGC TGGGGACTCT GGGCAGAAGG TGGCCCTGGA
TGGCTGGGGA GATGTCGAGC TGCATCAGTA GCTCTCTCGT CCCTGGGGCC                600
ACATAGGCCC TGAGGCATGT CACCACAAGT CCCCACTGCC AGCTGAGTCC
AGGGTGCCAG GGCTGAGAGA GGAAGTGAAA TTTATGATGC TTTCTTTCTT                700
TTTCCTCAGT TTGTGGGAAA TCGACACCTG AAAAAGAGGT GAGATGAAAT
C
GAGAGAGTTA CTCCCAAATG TCCCTGACCA TTCCTTATAA TTGCCTAATG                800
CTCAGATCCC CTGGAATCAT CCTTCACTTT CCGGGGGCTC GCCTCATTCC
CTCTAAGTCC CAACCCCCAC GTAGAATAAA GAGGGCCGGG GCTGGTTTTC                900
GCTGCCGCAC TAATCTGCGC CACCTTCTCT CTTTCAGGGG GAGCTTGAAG
GAACTACTAC TAAGCCCCTG GCCCCAAACC CAAGCTTCAG TCCCACTCCA               1000
GGGTTCACCC CCACCCTGGG CTTCAGTCCC GTGCCCAGTT CCACCTTCAC
CTCCAGCTCC ACCTATACCC CCGGTGACTG TCCCAACTTT GCGGCTCCCC               1100
GCAGAGAGGT GGCACCACCC TATCAGGGGG CTGACCCCAT CCTTGCGACA
GCCCTCGCCT CCGACCCCAT CCCCAACCCC CTTCAGAAGT GGGAGGACAG               1200
CGCCCACAAG CCACAGAGCC TAGACAGTGC GTTTCTCCCG CGGCTGGAGA
CGAGGAGGCT GGGGGAGGGC CGGGGGAGCG CGGGAGGCGT CAGAGGGGAC               1300
CACGAGAGGC GGAGGGCGCG GGATGCGGGG CGGGGCCTGG GGTTGCCGCC
CGAGGCTCAC CGGCCCGCGT CCCCGCAGCT GATGACCCCG CGACGCTGTA               1400
CGCCGTGGTG GAGAACGTGC CCCGTTGCG CTGGAAGGAA TTCGTCGCGC
GCCTAGGGCT GAGCGACCAC GAGATCGATC GGCTGGAGCT GCAGAACGGG               1500
CGCTGCCTGC GCGAGGCGCA ATACAGCATG CTGGCGACCT GGAGGCGGCG
CACGCCGCGG CGCGAGGCCA CGCTGGAGCT GCTGGGACGC GTGCTCCGCG               1600
ACATGGACCT GCTGGGCTGC CTGGAGGACA TCGAGGAGGC GCTTTGGGGC
CCCGCCGCCC TCCCGCCCGC GCCCAGTCTT CTCAGATGAG GCTGCGCCCC               1700
TGCGGGCAGC TCTAAGGACC GTCCTGCGAG ATCGCCTTCC AACCCCACTT
TTTTCTGGAA AGGAGGGGTC CTGCAGGGGC AAGCAGGAGC TAGCAGCCGC               1800
CTACTTGGTG CTAACCCCTC GATGTACATA GCTTTTCTCA GCTGCCTGCG
CGCCGCCGAC AGTCAGCGCT GTGCGCGCGG AGAGAGGTGC GCCGTGGGCT               1900
CAAGAGCCTG AGTGGGTGGT TTGCGAGGAT GAGGGACGCT ATGCCTCATG
CCCGTTTTGG GTGTCCTCAC CAGCAAGGCT GCTCGGGGGC CCCTGGTTCG               2000
TCCCTGAGCC TTTTTCACAG TGCATAAGCA GTTTTTTTTG TTTTTGTTTT
```

FIGURE 3A

GTTTTGTTTT GTTTTTAAAT CAATCATGTT ACACTAATAG AAACTTGGCA 2100
CTCCTGTGCC CTCTGCCTGG ACAAGCACAT AGCAAGCTGA ACTGTCCTAA
GGCAGGGGCG AGCACGGAAC AATGGGGCCT TCAGCTGGAG CTGTGGACTT 2200
TTGTACATAC ACTAAAATTC TGAAGTTAAA GCTCTGCTCT TGGAGACAGT
GGCT

FIGURE 3B

POLYMORPHISMS IN THE CODING SEQUENCE FOR THE TUMOR NECROSIS FACTOR RECEPTOR 1 (TNFR1)

```
ATGGGCCTCT CCACCGTGCC TGACCTGCTG CTGCCACTGG TGCTCCTGGA
                                     G
GCTGTTGGTG GGAATATACC CCTCAGGGGT TATTGGACTG GTCCCTCACC      100
TAGGGGACAG GGAGAAGAGA GATAGTGTGT GTCCCCAAGG AAAATATATC
CACCCTCAAA ATAATTCGAT TTGCTGTACC AAGTGCCACA AAGGAACCTA      200
CTTGTACAAT GACTGTCCAG GCCCGGGGCA GGATACGGAC TGCAGGGAGT
                          T
GTGAGAGCGG CTCCTTCACC GCTTCAGAAA ACGACCTCAG ACACTGCCTC      300
                T
AGCTGCTCCA AATGCCGAAA GGAAATGGGT CAGGTGGAGA TCTCTTCTTG
CACAGTGGAC CGGGACACCG TGTGTGGCTG CAGGAAGAAC CAGTACCGGC      400
          A
ATTATTGGAG TGAAAACCTT TTCCAGTGCT TCAATTGCAG CCTCTGCCTC
  C
AATGGGACCG TGCACCTCTC CTGCCAGGAG AAACAGAACA CCGTGTGCAC      500
CTGCCATGCA GGTTTCTTTC TAAGAGAAAA CGAGTGTGTC TCCTGTAGTA
ACTGTAAGAA AAGCCTGGAG TGCACGAAGT TGTGCCTACC CCAGATTGAG      600
AATGTTAAGG GCACTGAGGA CTCAGGCACC ACAGTGCTGT TGCCCCTGGT
CATTTTCTTT GGTCTTTGCC TTTTATCCCT CCTCTTCATT GGTTTAATGT      700
ATCGCTACCA ACGGTGGAAG TCCAAGCTCT ACTCCATTGT TTGTGGGAAA
TCGACACCTG AAAAAGAGGG GGAGCTTGAA GGAACTACTA CTAAGCCCCT      800
GGCCCCAAAC CCAAGCTTCA GTCCCACTCC AGGCTTCACC CCCACCCTGG
GCTTCAGTCC CGTGCCCAGT TCCACCTTCA CCTCCAGCTC CACCTATACC      900
CCCGGTGACT GTCCCAACTT TGCGGCTCCC CGCAGAGAGG TGGCACCACC
CTATCAGGGG GCTGACCCCA TCCTTGCGAC AGCCCTCGCC TCCGACCCCA     1000
TCCCCAACCC CCTTCAGAAG TGGGAGGACA GCGCCCACAA GCCACAGAGC
CTAGACACTG ATGACCCCGC GACGCTGTAC GCCGTGGTGG AGAACGTGCC     1100
CCCGTTGCGC TGGAAGGAAT TCGTGCGGCG CCTAGGGCTG AGCGACCACG
AGATCGATCG GCTGGAGCTG CAGAACGGGC GCTGCCTGCG CGAGGCGCAA     1200
TACAGCATGC TGGCGACCTG GAGGCGGCGC ACGCCGCGGC GCGAGGCCAC
GCTGGAGCTG CTGGGACGCG TGCTCCGCGA CATGGACCTG CTGGGCTGCC     1300
TGGAGGACAT CGAGGAGGCG CTTTGCGGCC CCGCCGCCCT CCCGCCCGCG
CCCAGTCTTC TCAGATGA
```

FIGURE 4

ISOFORMS OF TUMOR NECROSIS FACTOR RECEPTOR 1 (TNFR1)

```
MGLSTVPDLL LPLVLLELLV GIYPSGVIGL VPHLADREKR DSVCPQGKYI
HPQNNSICCT KCHKGTYLYN DCPGPGQDTD CRECESGSFT ASENHLRHCL        100
                          L                   M
SCSKCRKEMG QVEISSCTVD RDTVCGCRKN QYRHYWSENL FQCFNCSLCL
                          Q                   H
NGTVHLSCQE KQNTVCTCHA GFFLRENECV SCSNCKKSLE CTKLCLPQIE        200
NVKGTEDSGT TVLLPLVIFF GLCLLSLLFI GLMYRYQRWK SKLYSIVCGK
STPEKEGELE GTTTKPLAPN PSFSPTPGFT PTLGFSPVPS STFTSSSTYT        300
PGDCPNFAAP RREVAPPYQG ADPILATALA SDPIPNPLQK WEDSAHKPQS
LDTDDPATLY AVVENVPPLR WKEFVRRLGL SDHEIDRLEL QNGRCLREAQ        400
YSMLATWRRR TPRREATLEL LGRVLRDMDL LGCLEDIEEA LCGPAALPPA
PSLLR                                                         455
```

FIGURE 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/04606

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C07H 21/02; C12Q 1/68; C12P 21/00; A01N 43/04; G06F 17/00
US CL : 536/23.1; 435/6; 800/4; 514/44; 345/418
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 435/6; 800/4; 514/44; 345/418

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LUCKENBACH et al. Restriction fragment length polymorphism: Molecular weight analysis and calculation with a scanner-based computer system. Electrophoresis. February 1994, Vol. 15, No. 2, pages 149-152, especially page 151, Figure 3.	24
A	STUBER et al. A genomic polymorphism within the tumor necrosis factor locus influences plasma tumor necrosis factor-alpha concentrations and outcome of patients with severe sepsis. Critical Care Medicine. December 1996, Vol. 24, No. 3, pages 381-384, see entire reference.	16-23
A	EP 0 606 869 A1 (YEDA RESEARCH AND DEVELOPMENT COMPANY, LTD.) 20 July 1994, see entire reference, especially sequence 3.	1-3, 6, 7

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 05 JUNE 2000	Date of mailing of the international search report 11 JUL 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Cynthia Wilder Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/04606

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 93/19777 A1 (IMMUNEX CORPORATION) 14 October 1993, see entire reference.	1-23
A	US 5,491,075 A (DESNICK et al.) 13 February 1996, see entire reference, especially sequence 13.	15 and 16

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/04606

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

Sequence search on all databases (STIC-biotech/Chemlib), medline biotechds, embase, scisearch, biosis, cancerlit.
Search terms: TNFR1, tumor necrosis factor, tumor necrosis factor receptor, polymorphism, TNG gene, genotyping, genotyp?, polymorphism, polymorphic site, haplotyp?, hapotyping, ps1-ps12, reference, population, mutation, cpu, central processiong unit, communication, genome, anthology, drug and screening, binding and activity, polymorphic variant, computer